



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2013

Das Wuchern der Rhizome: Linguistische Diskursanalyse und Data-driven Turn

Scharloth, Joachim ; Eugster, David ; Bubenhofer, Noah

DOI: https://doi.org/10.1007/978-3-531-18910-9_11

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-111264>

Book Section

Accepted Version

Originally published at:

Scharloth, Joachim; Eugster, David; Bubenhofer, Noah (2013). Das Wuchern der Rhizome: Linguistische Diskursanalyse und Data-driven Turn. In: Busse, Dietrich; Teubert, Wolfgang. Linguistische Diskursanalyse: Neue Perspektiven. Wiesbaden: Springer, 345-380.

DOI: https://doi.org/10.1007/978-3-531-18910-9_11

Das Wuchern der Rhizome. Linguistische Diskursanalyse und Data-driven Turn

Joachim Scharloth (Dresden), David Eugster (Zürich), Noah Bubenhofer (Dresden)

1. Die linguistische Diskursanalyse am Scheideweg

1.1 Zum Stand der linguistischen Diskursanalyse: Ein Loblied

Die Geschichte der linguistischen Diskursanalyse in den nunmehr 25 Jahren seit dem Erscheinen von Dietrich Busses Dissertation „Historische Semantik. Analyse eines Programms“ (Busse 1987) ist eine Erfolgsgeschichte.

(a) Was 1987 noch als Kritik an einer historiographisch interessierten historischen Semantik konzipiert war, hat sich längst von der Konzeption eines methodologischen Gegenentwurfs zur Begriffsgeschichte emanzipiert und ist zu einer eigenständigen Teildisziplin der Linguistik mit unterschiedlichen Schulen und einem differenzierten Methodenrepertoire herangereift. Musste man sich von Fachkollegen Ende der 1990er Jahre noch fragen lassen, ob diskursanalytisches Arbeiten überhaupt etwas mit Linguistik zu tun habe, gelten heute solche Fragen nur noch als Zeichen der Ignoranz und wissenschaftlichen Fortschrittsfeindlichkeit des Fragestellers.

(b) In wissenschaftssoziologischer Perspektive ist die akademische Etablierung der linguistischen Diskursanalyse an verschiedenen Entwicklungen ablesbar: Regelmäßig werden auf den wichtigsten nationalen und internationalen Fachtagungen Panels mit diskursanalytischem Schwerpunkt organisiert. Wie es sich für eine Teildisziplin der Linguistik gehört, finden Lehrbücher mit einführendem Charakter Eingang in die Lehrbuchreihen wichtiger Wissenschaftsverlage.¹ Es entstehen zahlreiche diskursanalytisch orientierte Dissertationen und Sammelbände, die in angesehenen Buchreihen oder gar in Buchreihen mit diskursanalytischem Schwerpunkt veröffentlicht werden.² Die akademische Etablierung der linguistischen Diskursanalyse ist auch daran abzulesen, dass ihre Protagonisten inzwischen mehrere Professuren besetzen und zwar nicht trotz, sondern gerade wegen ihres Engagements für die junge Teildisziplin.

(c) Die linguistische Diskursanalyse ist gut vernetzt mit anderen sozial- und kulturwissenschaftlichen Disziplinen, die ebenfalls mit diskursanalytischen Methoden arbeiten, insbesondere den Politik- und Geschichtswissenschaften, der Soziologie und Sozialgeographie³, und fungiert in diesen Disziplinen vor allem als methodologische Inspirationsquelle.

Die Geschichte der linguistischen Diskursanalyse ist also eine Erfolgsgeschichte.

1.2 Zum Stand der linguistischen Diskursanalyse: Eine Invektive

Die linguistische Diskursanalyse hat ihre akademische Etablierung mit dem Verlust an gesellschaftsanalytischer Potenz und gesellschaftlicher Relevanz erkaufte.

(a) Die linguistische Diskursanalyse begnügt sich damit, mit dem Anspruch der linguistischen Disziplinierung und der Verwissenschaftlichung immer komplexere

¹ Vgl. die exzellente Monographie von Spitzmüller / Warnke 2011.

² Vgl. die Reihe „Sprache und Wissen“, herausgegeben von Ekkehard Felder.

³ Vgl. etwa das Forschungsnetzwerk „DiskursNetz“ und seine Online-Präsenz www.diskursanalyse.net (7.8.2012).

Kataloge mit Analysekategorien und Methoden zu entwerfen⁴, deren Umsetzung immer mehr Ressourcen verschlingt, deren heuristischer Mehrwert aber nicht belegt ist. Diskurslinguistinnen und -linguisten ernten auf interdisziplinären Konferenzen angesichts der geringen Originalität ihrer mit großem Methodenrepertoire gewonnenen Erkenntnisse häufig ein Achselzucken von ihren Kolleginnen und Kollegen aus den anderen Sozial- und Kulturwissenschaften. Während geschichtswissenschaftliche Diskursanalysen in Feuilletons besprochen werden und die gesellschaftlichen Debatten befeuern⁵ fristen diskurslinguistische Studien, die Arbeiten der Autoren dieses Beitrags eingeschlossen (Scharloth 2005, Bubenhofer 2009), ein akademisches Nischendasein.

(b) Zwar suggerieren die Kategorienkataloge und Methodenrepertoire wissenschaftliche Objektivität, eine präzise Bestimmung dessen, was Variation in den einzelnen Kategorien über geistes-, ideologie- oder wissensgeschichtliche Entwicklungen aussagt, eine Validierung ihrer Messinstrumente und eine Beurteilung ihrer Reliabilität ist die linguistische Diskursanalyse aber letztlich schuldig geblieben.

(c) Wegen des methodischen Korsetts, das ihnen aufgezwungen wird, sind Diskurslinguisten immer gezwungen, ihre Datenbasis „aus forschungspraktischen Gründen“ (Busse / Teubert 1994: 14) einzuschränken. In einer Zeit, in der gesellschaftliche Debatten fast ausschließlich in digitalen Medien verhandelt werden und Texte daher leicht zugänglich sind, erscheint uns diese Selektivität nicht mehr zeitgemäß.

(d) Die traditionelle linguistische Diskursanalyse ist (teilweise auch aus den genannten forschungspraktischen Gründen) dem Dogma der thematischen, gegenstands- oder wissenskomplexspezifischen Bestimmung von Diskursen verhaftet⁶, obwohl die grundlegenden Schemata der Wissensstrukturierung bzw. -generierung, die sie aufzudecken hofft, gerade nicht themengebunden sind.

(e) Die traditionelle linguistische Diskursanalyse, obwohl schon früh durch Wolfgang Teubert an die Korpuslinguistik angeschlossen⁷, hinkt den Entwicklungen in der Korpus- und Computerlinguistik hinterher. Multivariate Verfahren und maschinelles Lernen werden von ihr in der Textanalyse ebenso selten genutzt wie Methoden der automatischen Erkennung linguistischer Kategorien oder von der Computerlinguistik erarbeitete Ressourcen wie semantische Taxonomien.

Während Diskurslinguistinnen und -linguisten noch immer mehrere Jahre lang Texte hermeneutisch interpretieren oder dekonstruieren und manuell kodieren, arbeiten in den Forschungsabteilungen von Suchmaschinenriesen, der Webmonitoringdienste und der Webtrendanalysten, aber auch in staatlich geförderten Projekten im Bereich der sogenannten Sicherheitsinformatik Computerlinguisten und Informatiker an Methoden des Data Mining, die in der Lage sind, Veränderungen in der semantischen Matrix in Echtzeit abzubilden. Die linguistische Diskursanalyse steht an einem Scheideweg. Wenn sie sich nicht bemüht, Anschluss an die sprachtechnologischen Entwicklungen zu finden, dann wird Spitzenforschung im Bereich Sprachanalyse als Gesellschaftsanalyse nicht mehr an Universitäten stattfinden, sondern in privatwirtschaftlichen Unternehmen. Es ist zu befürchten, dass damit auch das kritische Potenzial dieser Forschung verloren

⁴ Vgl. exemplarisch Busse 1987, Spitzmüller / Warnke 2008.

⁵ Verwiesen sei hier exemplarisch auf die Monographien von Philipp Sarasin (2001, 2004) und den Sammelband von Sarasin et al. (2007).

⁶ So schreiben Busse / Teubert (1994: 14): „Zu einem Diskurs gehören alle Texte, die sich mit einem als Forschungsgegenstand gewählten Gegenstand, Thema, Wissenskomplex oder Konzept befassen“. Und Spitzmüller / Warnke (2011: 121f.) setzen die Notwendigkeit einer Themenorientierung bzw. eines Gegenstandsbezuges selbstverständlich voraus.

⁷ Vgl. Busse / Teubert 1994: 14-18 und neuerlich die konzise Einführung Teubert / Čermáková (2007).

geht. Die Marginalisierung der traditionellen akademischen Diskurslinguistik wird dann weiter fortschreiten.

2. Linguistische Diskursanalyse und Data-driven Turn

Die Digitale Revolution bedeutet für die Linguistik eine radikale Neuorientierung. Durch Digitalisierung steigt nicht nur die Verfügbarkeit sprachlicher Daten, für die Linguistik von besonderer Bedeutung ist vielmehr, dass sprachliche Daten durch ihre Digitalisierung auf neue Weise verarbeitbar werden. Mit der Steigerung der Speicher- und Rechenkapazitäten von Computern ist es zudem möglich geworden, große Datenmengen nicht nur zu befragen, sondern induktiv Strukturen in den Daten zu entdecken. Dies möchten wir anhand der Korpuslinguistik illustrieren.

2.1 Corpus Based vs. Corpus Driven

Zwar gilt die Korpuslinguistik als keine Subdisziplin der Linguistik sondern als eine Methode, korpuslinguistisches Arbeiten folgt aber einer eigenen Logik und generiert einen Denkstil, der viele Bereiche der Sprachwissenschaft nachhaltig verändert. Am ehesten der Vorstellung von Korpuslinguistik als einer Methode entspricht das corpus-based-Paradigma innerhalb der Korpuslinguistik. Digitale Korpora dienen demnach der Überprüfung von Forschungshypothesen. Die Hypothesen, die unabhängig von der Analyse des Korpus entwickelt wurden, formulieren Annahmen über interpretative Konstrukte, die mittels bereits bewährter interpretativer linguistischer Analysekatgeorien an einem Korpus überprüft werden sollen.

Diesem deduktiven Vorgehen steht die Möglichkeit eines induktiven Vorgehens zur Seite, das die Grundlage des corpus-driven-Paradigma bildet. Digitale Korpora sind hier nicht nur „Belegsammlungen oder Zettelkästen in elektronischer Form“, sondern ermöglichen eine eigene „korpuslinguistische Perspektive“ (Perkuhn / Belica 2006: 2). Statt eine Hypothese mit vorher festgelegten Analysekatgeorien zu überprüfen, werden in einem Korpus sämtliche Muster berechnet, die sich bei der Anwendung vorher festgelegter Algorithmen ergeben. Diese Muster werden im Anschluss kategorisiert. Damit geraten häufig Evidenzen in den Fokus, die entweder quer zu den vorher existierenden Erwartungen stehen und die Grundlage für neue Hypothesen sind, oder im besten Fall sogar solche Evidenzen, die die Bildung neuer interpretativer linguistischer Analysekatgeorien nahelegen.⁸

Es ist dieses Potenzial datengeleiteter Analysen, das es erlaubt, der Korpuslinguistik über eine empirische Methode hinaus den Status eines Denkstils zuzuschreiben. Denn das Ernstnehmen empirischer Widerständigkeiten, die sich mit traditionellen linguistischen Katgeorien nicht beschreiben lassen, birgt die Möglichkeit eines neuen Zugangs zu Sprache und den Katgeorien ihrer Beschreibung.⁹

⁸ Vgl. für eine ausführliche Diskussion von corpus-based und corpus-driven-Ansätzen Bubenhofer (2009: 99ff.). Das corpus-driven-Paradigma war bereits bei Sinclair (1991) angedacht und wird bei Tognini-Bonelli (2001: 65ff.) explizit gemacht. Im deutschsprachigen Raum verfolgen z.B. Arbeiten von Kathrin Steyer (Steyer 2004, Steyer / Lauer 2007, Steyer / Brunner 2009) dieses Paradigma, die auf Konzepten und Methoden von Cyril Belica, Rainer Perkuhn, Holger Keibel und Marc Kupietz fußen (Belica 1996, Belica 2001, Perkuhn / Belica 2006).

⁹ Zwar verzichtet das datengeleitete Paradigma auf das Formulieren von Hypothesen und auf eine Festlegung auf bestimmte Analysekatgeorien, es ist jedoch offensichtlich, dass auch beim datengeleiteten Verfahren vorgängiges Wissen in den Forschungsprozess einfließt, und zwar: (1) durch die Wahl der

2.2 Der Data-driven Turn und seine Folgen für die linguistische Diskursanalyse

An diesem Wandel im Umgang mit Daten partizipiert nicht nur die Korpuslinguistik. Auch in anderen Kultur- und Sozialwissenschaften werden datengeleitete Methoden zur Generierung neuer Fragestellungen eingesetzt. Die Fortschritte in der Digitalen Bildverarbeitung ermöglichen es, typische Konfigurationen in visuellen Darstellungen datengeleitet zu ermitteln; der Wandel von Musik wird anhand von strukturentdeckenden Verfahren über große Mengen digitaler Musikstücke berechenbar; die Geschichtswissenschaft erfindet sich unter dem Label „Data Driven History“ neu; in der Soziologie werden Daten aus sozialen Netzwerken dazu benutzt, die lebensstilsspezifische Gliederung sozialer Gemeinschaften aufzudecken; und in der Kunstgeschichte lassen sich auf der Basis von bildspezifischer Figurationen Sujets oder ganze Kunststile berechnen. Diese Entwicklungen haben das Potenzial, die Kultur- und Sozialwissenschaften nachhaltig zu verändern, weshalb wir von einem Data-driven Turn sprechen wollen.

Den skizzierten Ansätzen ist gemeinsam, dass sie auf vorgängige Hypothesen verzichten und mit Datenmengen arbeiten, die so umfangreich sind, dass sie von keinem wissenschaftlichen Individuum mehr in eine Gesamtschau gebracht werden können. Die Forschergruppe *semtracks*, der die Autoren dieses Beitrags angehören, arbeitet seit einigen Jahren daran, Methoden der von uns mitentwickelten datengeleiteten Korpuspragmatik auch für die linguistische Diskursanalyse fruchtbar zu machen.¹⁰ Ziel einer datengeleiteten linguistischen Diskursanalyse kann es freilich nicht sein, die Analysekategorien der traditionellen Diskursanalyse korpuslinguistisch zu operationalisieren oder gar Lektüren zu simulieren. Sie folgt vielmehr der Logik datengeleiteter Forschung und generiert Analysekategorien aus den Daten selbst.¹¹

Vor dem skizzierten Hintergrund (Verfügbarkeit großer Datenmengen, Data-driven Turn) halten wir zwei Dogmen der linguistischen Diskursanalyse für obsolet:

(1) Die Thematisität bzw. den Gegenstandsbezug von Diskursen: Eine forschungspraktische Beschränkung auf thematische, gegenstands- oder wissenskomplexspezifische Bestimmung von Diskursen¹² erscheint uns nicht geboten. Diskurse sollten vielmehr anhand strukturentdeckender Verfahren aus den Daten selbst berechnet werden.

Korpora, (2) hinsichtlich der Gestaltung der Algorithmen zur Musterberechnung, (3) bei der Festlegung dessen, was als linguistische Untersuchungseinheit (token) gelten soll, und (4) bei der Festlegung dessen, welche Einheitentypen eigentlich als potenzieller Bestandteil eines Musters aufgefasst werden sollen. Schließlich ist (5) auch das Kategorisieren der Daten im Anschluss an die Musterberechnung ein interpretativer Prozess, der zwar durch statistische Verfahren teilweise objektiviert werden kann; dennoch ist die Menge der Daten meist so umfangreich, dass eine weitere Reduzierung und Gewichtung im Sinne des Forschungsinteresses vorgenommen werden muss.

¹⁰ Vgl. insbesondere Bubenhofer 2009, Scharloth et al. 2010, Scharloth / Bubenhofer 2012, Bubenhofer / Scharloth in Druck.

¹¹ Wir gehen dennoch nicht so weit wie Chris Anderson, der Herausgeber des *Wired Magazine*, der schreibt: „the opportunity is great: The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.“ (Anderson 2008) Wir halten es für wichtig, dass die Ergebnisse datengeleiteter Analysen valide sind und nicht lediglich statistisch signifikant.

¹² Vgl. Abschnitt 1.2, Kritikpunkt (d).

(2) Die Forderung nach Lektüre bzw. der qualitativen Auswertung der Daten¹³: Die Datenmengen, mit denen nach unseren Vorstellungen im Rahmen datengeleiteter Analysen gearbeitet werden sollte, sind viel zu umfangreich, als dass sie noch durch Lektüre erschlossen, geschweige denn aufwändig kodiert werden könnten. Eine qualitative Deutung jedes einzelnen Textemplars scheidet daher von vornherein als Option aus. Die an die Adresse korpuslinguistischer Verfahren formulierte Kritik, mit der die Notwendigkeit qualitativen Arbeitens begründet wird, lautet, dass die Korpuslinguistik Daten quasi atomistisch, d.h. losgelöst von ihrem Kontext untersuche¹⁴, und nur durch die Lektüre ein kontextsensitives Deuten des Zeichengebrauchs möglich sei. Dieses Bild von Korpuslinguistik halten wir jedoch für überholt, denn die Analyse von Kookkurrenzen gehört zum Standardrepertoire der Korpuslinguistik.

Wir wollen in den folgenden Abschnitten illustrieren, wie eine datengeleitete linguistische Diskursanalyse aussehen könnte. Insbesondere möchten wir zeigen, dass mit Hilfe korpusgeleiteter Verfahren diskursive Umbrüche identifizierbar und analysierbar sind.¹⁵

3. Datengeleitete Frameanalyse am Beispiel des ZEIT-Archivs I: Identifizierung und Analyse von Umbruchzeiten

3.1 Zum Frame-Begriff

Im Folgenden wollen wir untersuchen, wie sich die Distribution und Vernetzung verschiedener Deutungsrahmen in der wöchentlich erscheinenden Zeitung DIE ZEIT in den Jahren von 1946 bis 2011 verändert haben. Hierbei beziehen wir uns auf den Begriff des Frames, wie er von Goffman (1974) expliziert wurde.¹⁶ Goffman geht es in seinem Buch „Frame Analysis“ darum, die „organization of experience“ (Goffman 1974: 13) theoretisch zu fassen, um Wissensstrukturen, „die es Menschen ermöglichen, ihre Erfahrungsdaten zu interpretieren“ (Ziem 2008: 14) Ein Frame ist für ihn dabei ein durch bestimmte Indikatoren hervorgerufenen Interpretationsschema, durch das eine Handlung erst einen Sinn erhält. So schreibt er über die grundlegenden Rahmen: „a primary framework is one that is seen as rendering what would otherwise be a meaningless aspect of the scene into something that is meaningful.“ (Goffman 1974: 21) Frames sind sozial geteilte und kulturspezifische Wissensstrukturen (vgl. Rettie 2004: 117).

Mit dem Begriff des Framing „können all diejenigen Prozesse bezeichnet werden, bei denen Deutungsmuster in der Informationsverarbeitung aktiviert werden.“ (Dahinden 2006: 28). Aus Sicht des öffentlichkeitstheoretischen Framingansatzes ist Framing der Prozess „by which a communication source, such as a news organization, defines and constructs a political issue or public controversy“ (Nelson / Oxley / Clawson 1997: 567).

¹³ Busse / Teubert (1994: 18) bezeichnen Diskursanalyse als „ein fortschreitend die Korpusbildung korrigierendes Lesen“. Spitzmüller / Warnke (2011: 39) fordern: „Die quantitative Organisation des Datenmaterials durch statistische Auswertungsmöglichkeiten der Korpuslinguistik sollte qualitative Analysen nicht ersetzen, sondern ergänzen.“

¹⁴ Vgl. etwa Spitzmüller / Warnke 2011: 38.

¹⁵ Beispielsweise im Widerspruch zur Kritik an unseren korpuslinguistischen Ansätzen in Spitzmüller / Warnke (2011: 39f).

¹⁶ Wir halten eine für korpusgeleitete Zwecke geeignete Operationalisierung framesemantischer Konzepte, wie sie etwa von Konerding (1993) und Ziem (2008) elaboriert wurden, für gegenwärtig nicht zu leisten.

In diesem wissenschaftlichen Kontext sind Frames „kollektive Deutungsmuster, in denen bestimmte Problemdefinitionen, Kausalzusammenhänge, Ansprüche, Begründungen und Wertorientierungen in einen mehr oder weniger konsistenten Zusammenhang gebracht werden, um Sachverhalte zu erklären, Kritik zu fundieren und Forderungen zu legitimieren“ (Neidhardt / Rucht 1993: 108).

Wir beschäftigen uns im vorliegenden Aufsatz also mit dem Wandel von Realitätskonstruktionen durch die Medien. Allerdings werden wir uns gemäß dem Programm einer datengeleiteten linguistischen Diskursanalyse nicht auf das Framing eines einzelnen Issues beschränken. Vielmehr wollen wir versuchen, die Konjunktoren von Frames nachzuzeichnen und den sich wandelnden Verbindungen der Frames untereinander auf die Spur zu kommen.

3.2 Das Untersuchungskorpus

Die folgenden Untersuchungen wurden anhand des Print-Archivs der Wochenzeitung DIE ZEIT vorgenommen. Das Korpus umfasste die kompletten Jahrgänge 1946-2011, wie sie auf zeit.de zu finden sind. Es umfasst insgesamt 271.439.149 laufende Wortformen. Die Anzahl der laufenden Wortformen je Jahrgang variiert dabei zwischen 1.051.351 (1946) und 6.520.382 (2005).

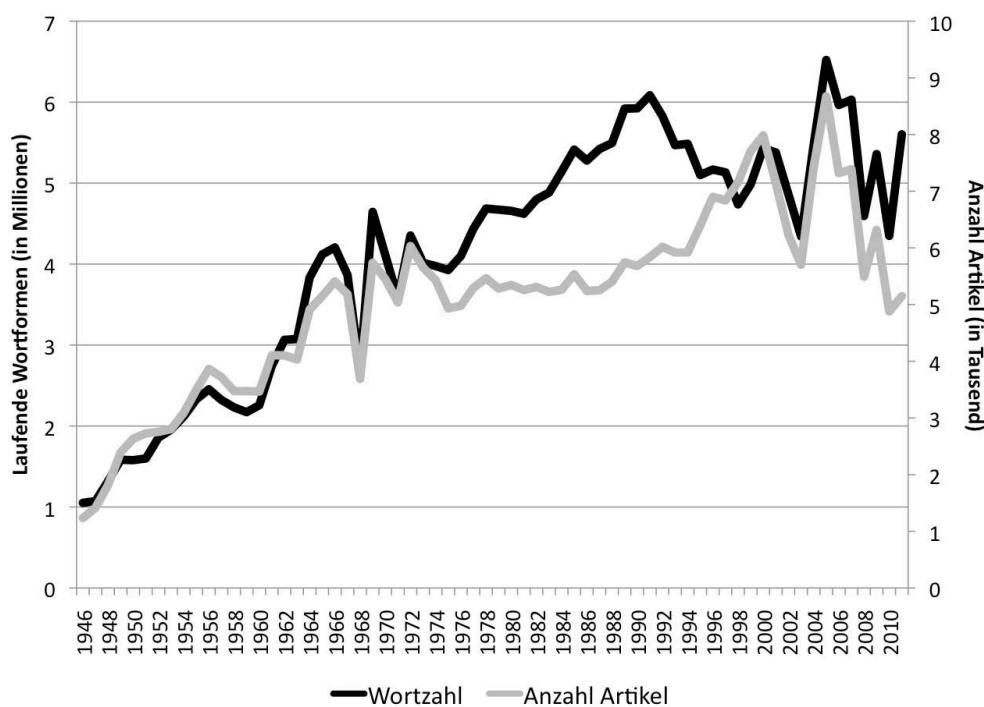


Abbildung 1: Entwicklung der Wortzahl und der Anzahl der Artikel im Print-Archiv der ZEIT, 1946-2011

Die Qualität des Korpus leidet darunter, dass erst für die Jahrgänge ab 1995 Texte aus internen elektronischen Datenbanken des Zeitverlags Gerd Bucerius zur Verfügung standen, die Texte der früheren Jahrgänge hingegen mittels OCR digitalisiert wurden und daher oft fehlerhaft sind. Das Korpus wurde zwar mit Metainformationen (neben dem Erscheinungsdatum der Texte die Autoren und das Ressort) annotiert, doch sind auch diese Informationen vor 1995 nur sehr unsystematisch und unvollständig

verfügbar. Insbesondere fehlen häufig Hinweise darauf, in welchem Ressort ein Artikel abgedruckt wurde. Ressortspezifische Untersuchungen über den gesamten Zeitraum waren daher nicht möglich.

Das Korpus wurde mit Hilfe des TreeTaggers (Schmid 1994) tokenisiert, mit Wortarten-Informationen annotiert und lemmatisiert. Beim verwendeten Tagset handelt es sich um das Stuttgart-Tübingen-Tagset (STTS) (Schill et al. 1995).

Die Beschränkung des Untersuchungskorpus auf sämtliche Texte einer Wochenzeitung begrenzt natürlich auch die Aussagekraft der Analysen, die mit diesem Korpus vorgenommen werden. Streng genommen untersuchen wir lediglich Wirklichkeitskonstruktionen, wie sie in den Texten der ZEIT vorgenommen wurden. Eine Ausweitung auf weitere Korpora wäre natürlich wünschenswert. Die folgenden Untersuchungen sollen aber vor allem exemplarischer Natur sein und das Potenzial der angewendeten Verfahren illustrieren.

3.3 Methode der Frameidentifikation

Für die vorliegende Studie sind wir davon ausgegangen, dass Frames in Texten durch das Auftreten bestimmter Lemmata identifizierbar sind. Wir haben Frames für die vorliegende Untersuchung mit Hilfe einer semantischen Taxonomie operationalisiert, dem „Deutschen Wortschatz nach Sachgruppen“ (Dornseiff 2004).¹⁷

Die Identifizierung des Auftretens eines Frames in einem Text wurde dabei an drei Kriterien gebunden:

- (1) Die relative Frequenz der vorkommenden Vertreter einer Sachgruppe im Verhältnis zur Wortzahl eines Textes.
- (2) Die absolute Frequenz der vorkommenden Vertreter einer Sachgruppe im Text.
- (3) Die Abdeckung der Lemmata einer Sachgruppe (d.h. wie viele der eine Sachgruppe konstituierenden Lemmata kommen im Text vor).

3.4 Zeitreihenanalyse

Je nach Forschungsinteresse kann die Untersuchung framespezifischer Zeitreihen interessante Ergebnisse liefern. Für die vorliegende Untersuchung haben wir jahresweise die relative Frequenz aller Frames berechnet. Um die Entwicklung der Frames vergleichbar zu machen, haben wir für die framespezifischen Zeitreihen eine Min-Max-Normalisierung vorgenommen. Um herauszufinden, welche Frames im untersuchten Zeitraum eine für unsere Forschungsinteressen relevante Verteilung aufweisen, haben wir die jahresweise Verteilung der Frames ebenso als Vektor aufgefasst wie die uns interessierende Näherungsverteilung. Mittels der Berechnung der euklidischen Distanz zwischen Näherungsverteilungsvektor und allen Framevektoren konnten wir so diejenigen Vektoren identifizieren, die der uns interessierenden Frameentwicklung am nächsten kamen. In den folgenden Analysen haben wir für die framespezifischen Zeitreihen jeweils mit dem gleitenden Durchschnitt (Periode 5) gearbeitet, um längerfristige Trends besser sichtbar zu machen.

¹⁷ Problematisch an der Arbeit mit einer Taxonomie für eine Analyse von Texten, die wie im vorliegenden Fall über einen längeren Zeitraum verteilt sind, ist ihre Statik. Der „Dornseiff“ enthält jedoch neben dem zentralen Wortschatz eine solche Breite an teils wenig gebäuchlichen Lemmata, dass wir ihn für hinreichend robust halten, mit seiner Hilfe die Zeitgeschichte zu untersuchen.

Die folgenden Abbildungen zeigen jene Frames, die im Untersuchungszeitraum näherungsweise stetig ab- (Abbildung 2) bzw. zunehmen (Abbildung 3), gemessen an einer stetigen Verteilung, die im Jahr 1948 den Wert 0 und im Jahr 2009 den Wert 1 hat.

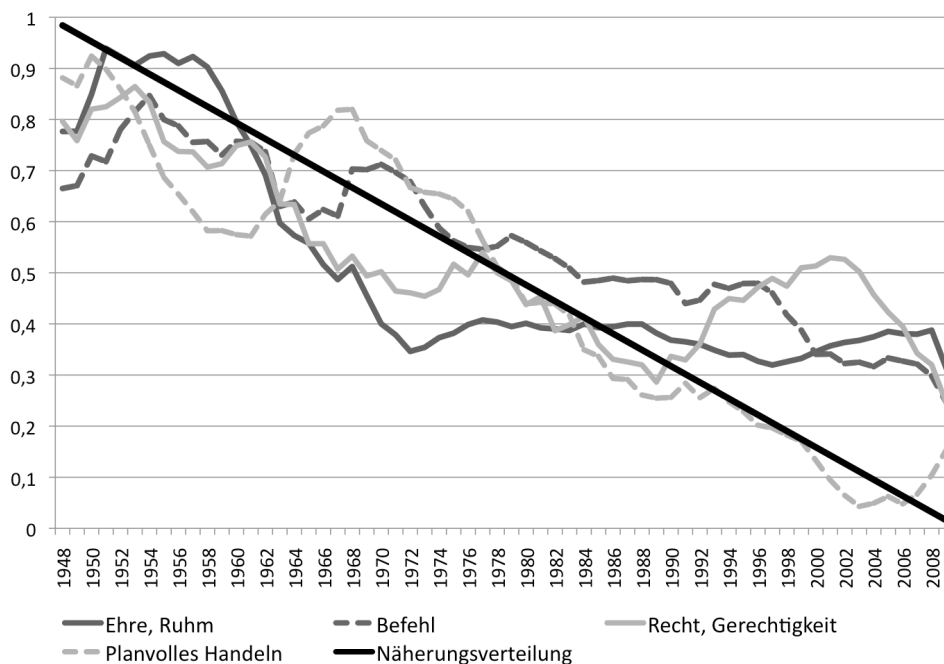


Abbildung 2: Frames, deren jahresweise relative Frequenz im Print-Archiv der ZEIT annähernd stetig abnimmt.

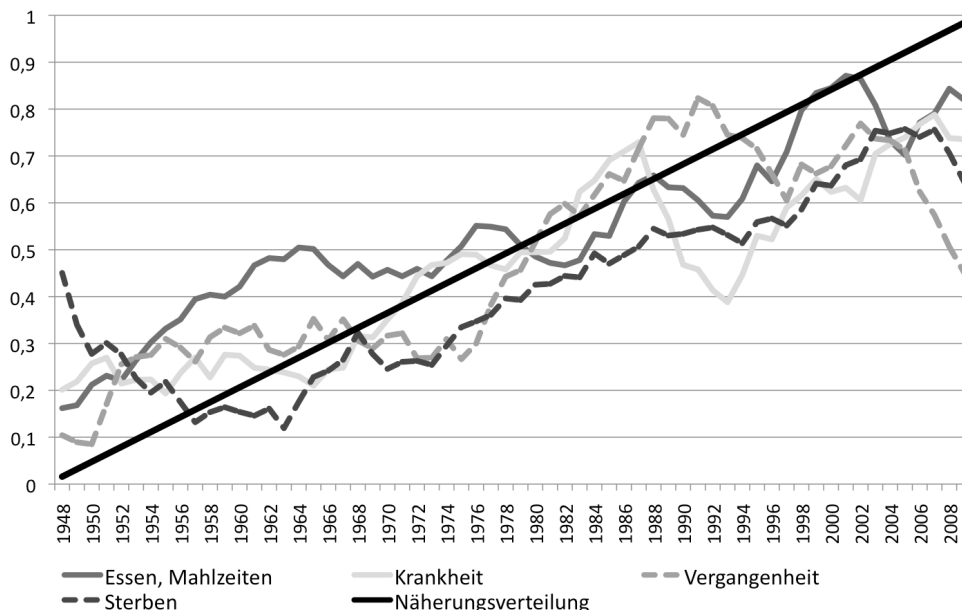


Abbildung 3: Frames, deren jahresweise relative Frequenz im Print-Archiv der ZEIT annähernd stetig zunimmt.

Es zeigt sich, dass unter den vier Frames mit der geringsten euklidischen Distanz zur stetig abnehmenden Näherungsverteilung mit „Ehre und Ruhm“ und „Befehl“ zwei Frames sind, die eher in einem traditionell autoritären Wertekanon von Bedeutung sind, der in der ZEIT offenbar stark an Bedeutung verloren hat. Die Tatsache, dass die Frames „Krankheit“ und „Sterben“ zu jenen vier Frames zählen, deren Auftreten stetig zunimmt,

lässt sich entweder als einer verstärkte Orientierung der ZEIT hin zu religiösen Themen (vgl. die Zunahme des „Übersinnlich“-Frames unten) oder der verstärkten Hinwendung zu Themen des Alterns deuten.

Zudem interessierte uns, welche Frames erst in den letzten Jahren eine besondere Konjunktur erlebten. Hierfür boten sich kubische (Abbildung 4) und exponentielle (Abbildung 5) Verteilung als Näherungsverteilungen an.

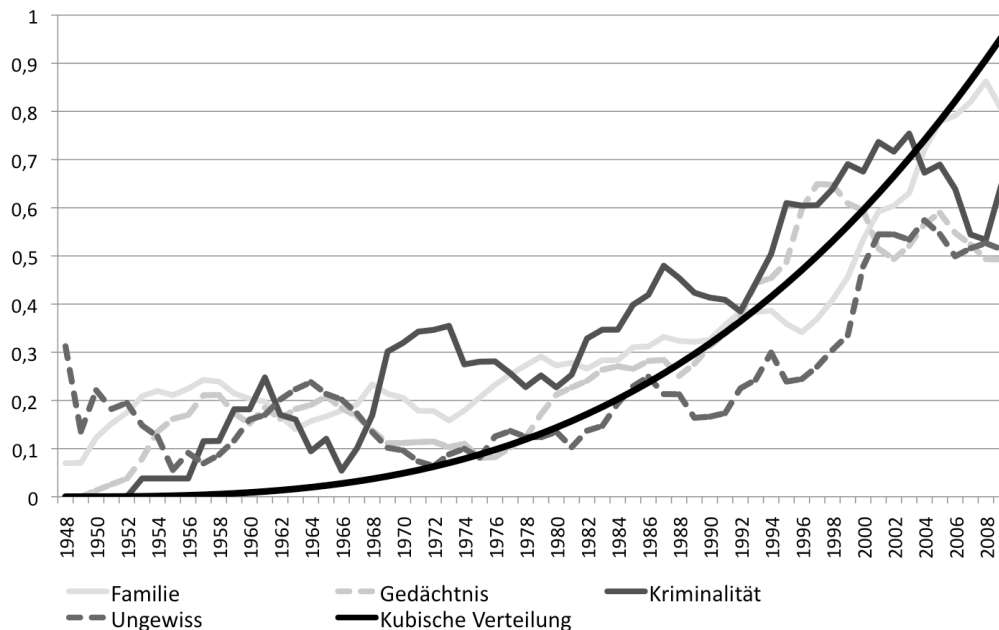


Abbildung 4: Frames, deren jahresweise relative Frequenz im Print-Archiv der ZEIT annähernd kubisch verteilt ist.

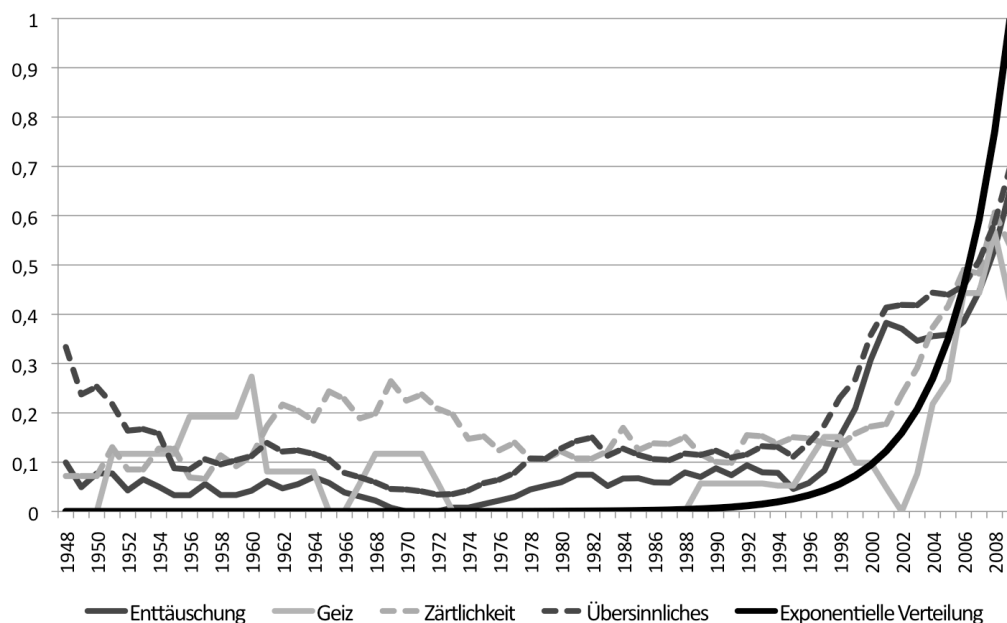


Abbildung 5: Frames, deren jahresweise relative Frequenz im Print-Archiv der ZEIT annähernd exponentiell verteilt ist.

Die Ergebnisse verdeutlichen, dass es keine im engeren Sinn politischen Frames sind, die in der Wochenzeitung DIE ZEIT in den letzten zehn bis zwanzig Jahren stark

zugewonnen haben. Vielmehr sind es – ein diskurslinguistisch durchaus interessanter Befund – Frames, die den breiteren gesellschaftlichen Kontext reflektieren. Doch liefert die Entwicklung einzelner Frames noch keine befriedigende Antwort auf die Frage, wann Umbrüche stattgefunden haben.

3.5 Sprachliche Umbrüche als Indikatoren zeitgeschichtlicher Umbrüche

Unter sprachlichen Umbrüchen verstehen wir mit Heidrun Kämper den „Beginn von sprachlichen Veränderungen“, die durch „plötzliche gesellschaftliche oder politische Veränderungen“ induziert werden und Ursache für einen längerfristigen sprachlichen Wandel sein können (Kämper 2007: 431). Das Konzept der sprachlichen Umbruchgeschichte impliziert damit, dass Sprache auf das Engste mit Gesellschafts- und Kultur- und politischer Geschichte verwoben ist und ihre Analyse einen Beitrag zum vertieften Verständnis der Qualität historischer Veränderungen leisten kann. Statt wie Kämper (2007: 432) die Identifizierung von Umbruchzeiten aus der Zeitgeschichtsschreibung zu übernehmen und nach korrespondierenden sprachlichen Phänomenen zu suchen, haben wir Umbrüche datengeleitet berechnet. Eine erste Annäherung erfolgte mittels der Analyse von Frames, die Zeitbezüge reflektieren. Als Charakteristikum von Umbruchzeiten identifiziert Kämper nämlich die Tendenz, dass die vom Umbruch betroffenen Gesellschaften ihr Selbstbild im Medium zeitreflexiver Diskurse verhandeln (vgl. Kämper 2007: 429). Abbildung 6 zeigt die Entwicklung von Frames mit Zeitbezug im ZEIT-Archiv.

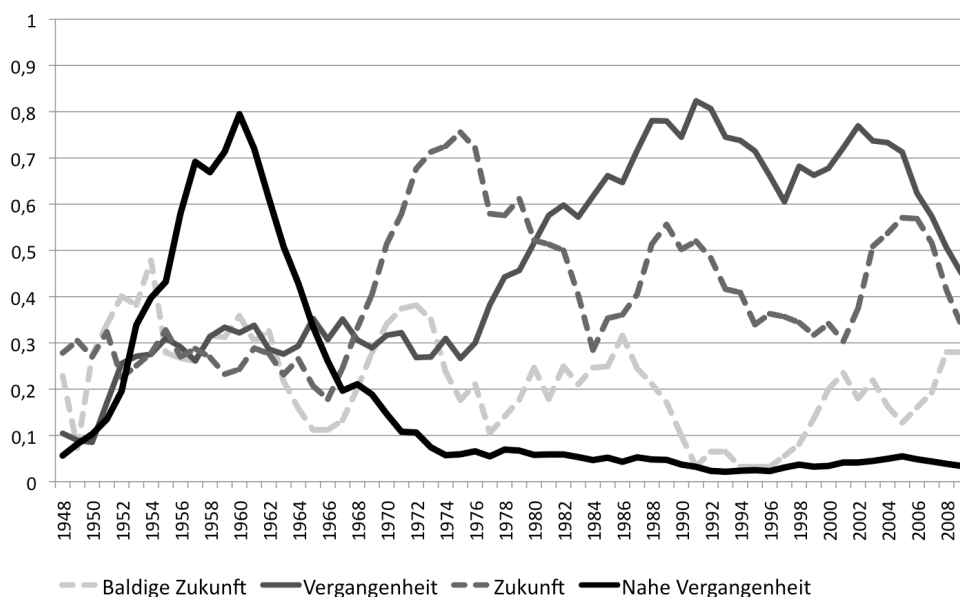


Abbildung 6: Jahresweise Verteilung von Zeitbezug-Frames im Printarchiv der ZEIT.

Anhand der Verteilung der Zeitbezug-Frames ließen sich durchaus Umbruchzeiten eingrenzen: etwa die späten 1950er / frühen 1960er Jahre, die 1970er Jahre, Anfang der 1990er Jahre und der Anfang der 2000er Jahre. Interessant ist, dass in den genannten Zeiträume jeweils andere Frames dominieren. Während der Frame „Nahe Vergangenheit“ insbesondere in den 1950er Jahren häufig in den Texten der ZEIT auftrat und der Frame „Vergangenheit“ Anfang der 1990er Jahre sein Maximum hat, hat der Frame „Zukunft“ in den 1970er Jahren besondere Konjunktur.

Zwar mag die Verteilung von Zeitbezug-Frames ein Indikator für das Vorliegen einer Umbruchzeit sein, für eine präzisere Bestimmung von Umbruchzeiten in der Geschichte der BRD sind wir von der Annahme ausgegangen, dass starke gesellschaftliche Veränderungen mit starken Veränderungen in der Verteilung aller Frames des Framehaushaltes korrelieren. Während Zeiten der Stabilität davon gekennzeichnet sind, dass nur geringfügige Verschiebungen im Framehaushalt zu beobachten sind, zeichnen sich Umbrüche dadurch aus, dass eine ein Teil der Frames stark abnimmt, während ein anderer Teil stark zunimmt. Deshalb halten wir die Summe der Veränderungen im Framehaushalt für ein geeignetes Maß für die Identifizierung von Umbruchzeiten.

Mathematisch haben wir diesen Ansatz auf zweierlei Weise umgesetzt: Zum einen haben wir für jedes Jahr den Betrag der Veränderung der relativen Frequenzen aller Frames im Vergleich zum Vorjahr berechnet und summiert; zum anderen haben wir die gleiche Berechnung mit den normalisierten Werte durchgeführt. Während bei der Berechnung der Differenzbeträge der relativen Frequenzen die hochfrequenten Frames ein höheres Gewicht haben, werden bei der Berechnung der Differenzbeträge der normalisierten Frequenzen alle Frames gleich gewichtet. Wie die folgende Grafik belegt, führen aber beide Berechnungsmethoden zu ähnlichen Ergebnissen:

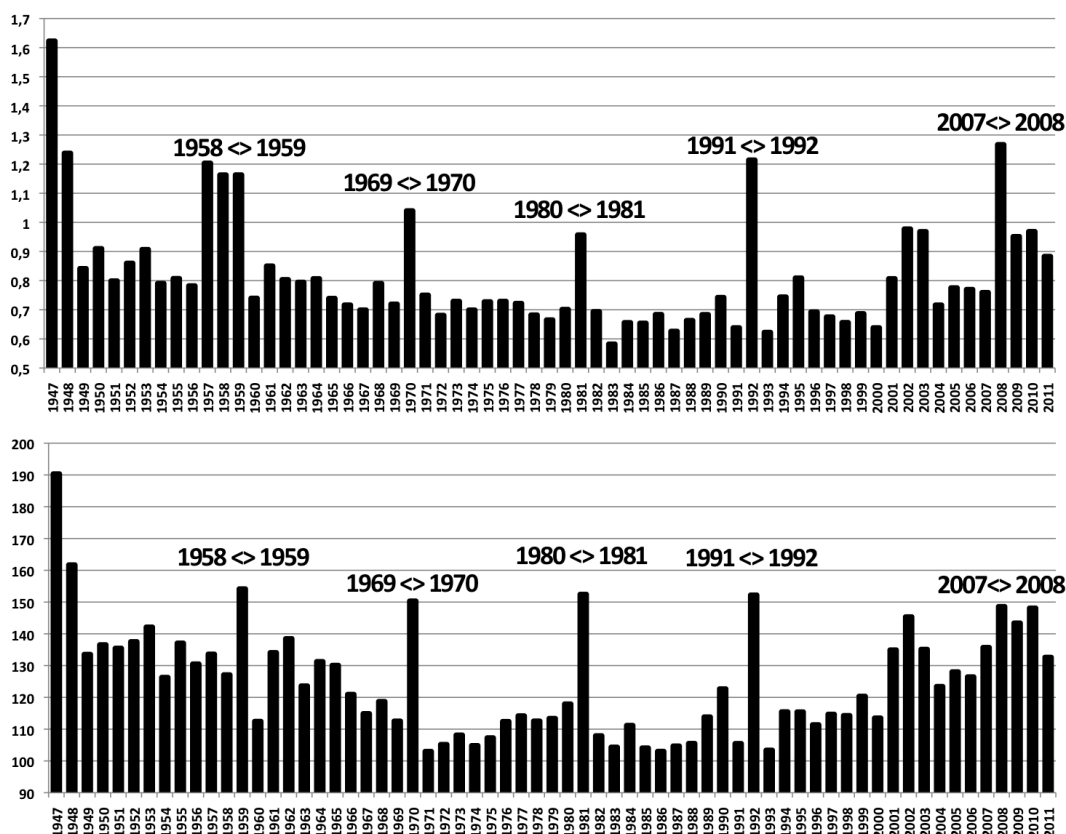


Abbildung 7: Jährliche Summen der Differenzbeträge aller Frames im Vergleich zum Vorjahr im Print-Archiv der ZEIT, 1946-2011. Oben: relative Frequenzen, Unten: normalisierte relative Frequenzen.

Die Grafiken zeigen, dass in den Jahren 1957-1959 (mit Schwerpunkt 1959), 1970, 1981, 1992 und 2008-2010 (mit Schwerpunkt 2008) besonders starke Veränderungen im Framehaushalt im Vergleich zu den Vorjahren zu beobachten sind. Auch die Jahre 2001-2003 können, wenn auch leicht abgeschwächt, als Jahre der Veränderung gelten.

Diese Zunahmen im Differenzbetrag deuten wir als Indikatoren für eine starke Veränderung in der semantischen Matrix und damit als Umbrüche im oben beschriebenen Sinn. Insbesondere bei den Umbrüchen von 1969/70, 1980/81 und 1991/92 sind in den folgenden Jahren nur vergleichsweise geringe Veränderungen zu beobachten, während nach den Umbruchjahren 1957-1959 und 2008-2010 eine allmähliche Verringerung der Variation zu beobachten ist.

Einige dieser anhand der Frameanalyse identifizierten Umbruchjahre lassen sich auf zeitgeschichtliche Ereignisse und Entwicklungen beziehen: der Umbruch von 1969/70 könnte als Folge der 68er-Bewegung gedeutet werden, die Veränderungen von 1991/92 als Nachwirkung der deutschen Einheit, die Variation in den Jahren 2001 bis 2003 als Effekt der Terroranschläge vom 11. September 2001 und die starken Veränderungen nach 2008 als Folge der Finanz- und Wirtschaftskrise. Bei den Umbruchjahren 1957 bis 1959 und 1980/81 ist es jedoch schwieriger, eine plausible zeithistorische Begründung zu finden. Können hier Wiederbewaffnung und Diskussion um die Ausstattung der Bundeswehr mit Atomwaffen, europäische Integration (1957-1959) und NATO-Doppelbeschluss und Friedensbewegung (1980/81) als Erklärung herangezogen werden? Im Grunde ist es müßig diese Fragen zu beantworten, solange nicht untersucht ist, welche Frames sich in den Umbruchzeiten besonders stark verändern.

Im Folgenden seien exemplarisch die Umbruchzeiten der späten 1950er und späten 1960er Jahre daraufhin untersucht, welche Frames im Verhältnis zu den Vor- und Folgejahren in ihnen besonders häufig verwendet wurden. Hier haben wir die euklidischen Distanzen zu einer als Vektor aufgefassten Gauß'schen Näherungsverteilung berechnet.

In den Jahren um 1959 (Abbildung 8) treten die Frames „Nahe Vergangenheit“, „Bergbau“, „Unternehmensgewinn“, „Sparsamkeit“, „Buchhaltung und Bilanzierung“ und „Gute Qualität“ besonders häufig auf. Die Frameanalyse legt also die Interpretation nahe, dass in der ZEIT ökonomischer Erfolg und bürgerliche Tugenden prägende Elemente der Adenauer-Ära waren. Dieser Befund ist zwar wenig originell, kann aber – wie auch die folgende Analyse zu den Jahren um 1969 – als Indiz für die Validität der Frameanalyse gelten.

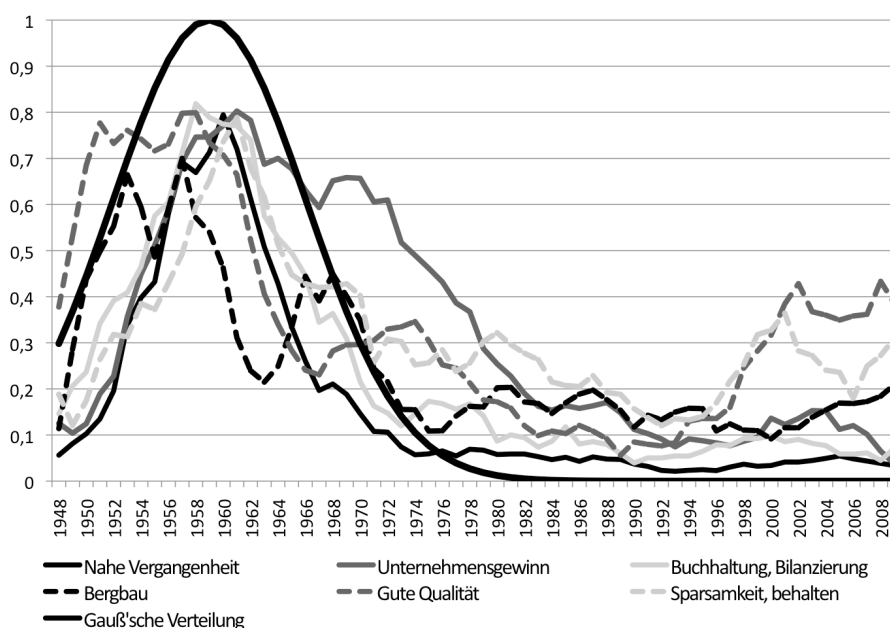


Abbildung 8: Frames im ZEIT-Archiv, die in den späten 1950er Jahren häufiger auftreten, als in den Vor- und Folgejahren

Im Unterschied zu den späten 1950er Jahren sind es in den späten 1960er und frühen 1970er Jahren die Frames „Regierung“, „Parlament“, „Friede“, „Vergnügen und Lachen“, „Widerstand“ und „Experiment“, die im Vergleich zu den Vor- und Folgejahren besonders häufig auftreten (Abbildung 9). Offenbar verweisen die Frames auf die politischen Konfrontationen und den gesellschaftlichen Aufbruch in den Jahren der Außerparlamentarischen Opposition. Interessant ist, dass der in der Forschung zur 1968er-Bewegung erst spät attestierte Hedonismus schon durch die Frameanalyse in den Texten der ZEIT sichtbar gemacht werden kann.

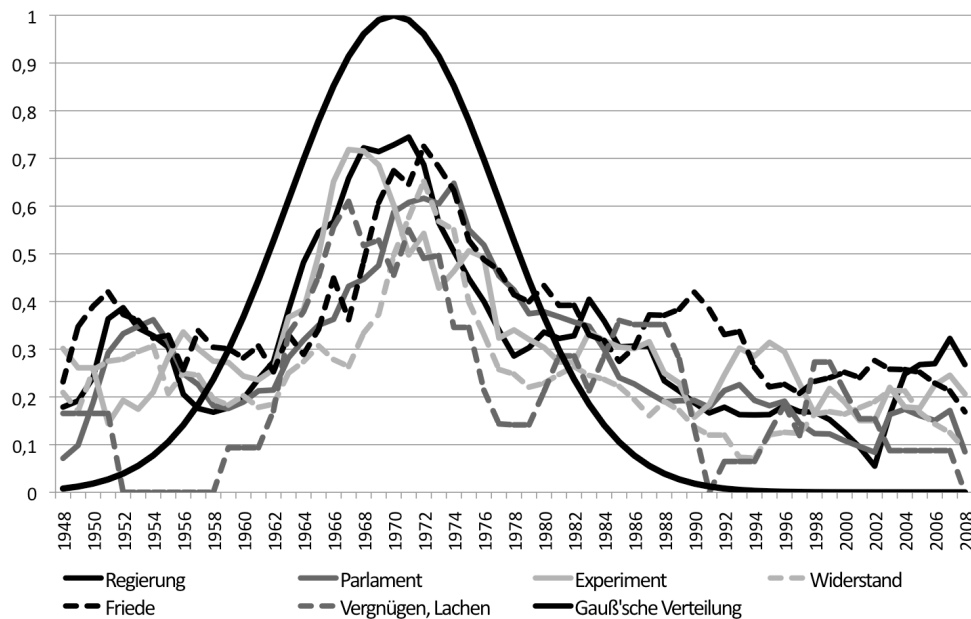


Abbildung 9: Frames im ZEIT-Archiv, die in den späten 1960er Jahren häufiger auftreten, als in den Vor- und Folgejahren

Wichtiger als die Entwicklung der Auftretenshäufigkeit einzelner Frames oder ihrer Kovariation aber ist die Frage, welche Frames *in einzelnen Texten* überzufällig häufig gemeinsam auftreten und ob hier im Laufe der Jahre, insbesondere während der Umbruchzeiten, Veränderungen zu beobachten sind.

4. Vom Wuchern der Rhizome: Das diskursive Geflecht als Kollokationsgraph

4.1 Diskurs als Wissensgeflecht

In der Diskurslinguistik ist regelmäßig vom Diskurs als „Geflecht“ die Rede. Der Diskurs wird als „Textgeflecht“, als „Textensemble“ beschrieben, das einzelne Texte verbinden (Hermanns 1995: 86f.). Die einzelnen Texte stehen zueinander in einem kommunikativen, intertextuellen Bezug, indem sie sich zitieren, kopieren, bzw. ähnliches thematisieren und gleiche oder ähnliche Äußerungen¹⁸ beinhalten. Das Textgeflecht des Diskurses basiert also letztlich auf in verschiedenen Texten verstreuten Äußerungen, die in der traditionellen linguistischen Diskursanalysen als einem bestimmten, thematisch gefassten Diskurs zugehörig beschrieben werden, weil sie sich ähnlich sind. Einen Diskurs so zu beschreiben, heißt die Kohärenz dieser verstreuten

¹⁸ Zur Differenz von Aussage und Äußerung vgl. Angermüller (2007).

Texte über die in ihnen gefundenen Aussagen zu behaupten. Doch es ist falsch anzunehmen, dass, fasst man den Diskursbegriff semantisch, diese Äußerungen nur einem Diskurs zugehören. Denn Äußerungen sind nie nur mit einem Textgeflecht verbunden, Texte haben stets etliche intertextuelle Anknüpfungen: „Aufgrund des Evokationspotentials der Sprachzeichen, die in der sprachlichen Formulierung, die die enoncé trägt, enthalten sind, ist jede einzelne Aussage (und eigentlich jedes einzelne in einem Text eingeführte epistemische Element) eingebettet in einen Kontext weiterer Wissensrahmen und Wissensrahmen-Geflechte.“ (Busse 2006: 27)

Will man einen Diskurs analysieren, nimmt man frühzeitig heuristische Beschneidungen vor. Gerade um den Diskurs als „Geflecht von thematisch zusammengehörigen Aussagen“ (Böke / Jung / Niehr / Wengeler 2000:12) herauszuarbeiten, wird die Verflechtung der einzelnen Äußerungen früh schon beschnitten. Aus arbeitstechnischen Gründen kann nicht jede Vernetzung, welche jede der vorgefundenen Äußerungen eingeht, berücksichtigt werden. Diskurslinguistik isoliert einzelne Bezüge aus Texten und ignoriert notwendigerweise andere. Am Ende präsentiert man eine aus dem Geflecht extrahierte intertextuelle Serie, einen diskursiven Strang. Das Geflecht, dem diese Äußerungen entnommen sind, bleibt unsichtbar. Diskurslinguistische Analysen wirken dadurch oft entdichtet. Trotz ihrer Prämissen suggerieren sie in ihrer Darstellung, die diskursive Organisation von Wissen sei in gewisser Form abgeschlossen betrachtbar. Dabei ist letztlich davon auszugehen, dass, nimmt man die oft verwendete Metapher des Geflechts ernst, Diskurse keine allzudeutlichen Ränder haben, keinen tabellarischen Ordnungen folgen, sondern vielmehr „wuchern“ (Bublitz 1999).

Bei Foucault finden sich keine Anregungen dazu, wie Diskurse als Geflecht dargestellt werden könnten. Fündig wird man im Denkkollektiv (Fleck), dem Foucault angehört hat. Gilles Deleuze und Felix Guattari haben in ihrem Einleitungstext von „Milles Plateaux“ – zu dessen Vorgänger „Anti-Ödipus“ Foucault die Einleitung verfasst hatte – den Begriff des Rhizoms entwickelt. In der Botanik beschreibt der Ausdruck ein Wurzelgeflecht, das unterirdisch wuchert und, wie z.B. eine gewundene Ingwerknolle, an jeder Stelle wieder ausschlagen kann. Deleuze und Guattari verwenden den Begriff des Rhizoms als Gegen-Denkfigur¹⁹ zu strukturalistischen Baum-Diagrammen, d.h. zu hierarchischen gerichteten Graphen. Das Rhizom bezeichnet für sie geflechtartige Beziehungen zwischen diversen kulturellen Praktiken, Machtstrukturen und Äußerungsarten (Deleuze/Guattari 1992: 12), beschreibt aber auch „kollektive Äußerungsgefüge“ (Deleuze/Guattari 1992: 16). Die Denkfigur des Rhizoms ermöglicht eine dehierarchisierte und dezentrierte Repräsentation kollektiven Wissens. Jeder Punkt, d.h. jede Äußerung, jede Praktik, ist potenziell über „Linien“ mit jedem anderen verknüpft, von jedem Punkt können Linien weg verweisen, um jeden Punkt können sich neue Punkte gruppieren. Das Rhizom hat keinen Ursprung, wohl aber Bündelungszentren und fixe Achsen, aus denen es wächst.

Streng genommen verweist ein Rhizom laut Deleuze und Guattari über die einzelnen Zeichen- und Praxisregime hinaus: Sprachliche Kettenglieder verweisen nicht notwendigerweise auf weitere sprachliche Kettenglieder, sondern können z.B. auch mit nicht-sprachlichen ökonomischen oder politischen Praktiken verknüpft werden (Deleuze/Guattari 1992: 16). In anderen Schriften haben Deleuze und Guattari aber durchaus Zugeständnisse gemacht, die für die Diskurslinguistik anschließbar sind, und analog funktionierende Strukturen etwas homogener beschrieben: In „Was ist Philosophie?“ nutzen Deleuze und Guattari analoge Ordnungsvorstellungen um das

¹⁹ Den Begriff der Metapher lehnen D und G ab: „Wir machen absolut keinen metaphorischen Gebrauch von diesen Begriffen... Wir meinen das so, wie wir es sagen: buchstäblich“ (Deleuze/Parnet 1980: 25).

Funktionieren philosophischer Begriffswelten zu beschreiben. Philosophie zu betreiben heißt nach Deleuze und Guattari, „Karten“ zu zeichnen, die beschreiben, in welchen „Regionen“ (Deleuze/Guattari 1996: 44) Begriffe sich mit welchem „Wucherungsgrad“ und mit welchen „Rückkoppelungen, Verbindungen“ zu anderen auf dieser Ebene sich bewegen (Deleuze/Guattari 1996: 48). Das Zeichnen der Karte vermag diesen Bewegungen und Verortungen „Konsistenz verleihen ohne irgend das Unendliche preiszugeben“ (Deleuze/Guattari 1996: 51). Auch in „Milles Plateaux“ wird diese Nähe des Rhizoms zur Karte angesprochen (Deleuze/Guattari 1996: 21). So ließe sich das Rhizom zur methodischen Schärfung enger gefasst auch als kartografische, d.h. grafische Repräsentation von epistemologisch relevanten Verknüpfungen zwischen Begriffen und auch zwischen Wörtern beschreiben (Deleuze/Guattari 1996: 29).

Deleuze und Guattari haben sich im ihnen eigenen Schreibhabitus kaum um die methodische Verwendung ihrer Terminologie geschert, ihnen ging es nicht zuletzt um die Beschreibung einer gegen den Strukturalismus gewendete Ontologie, die sich wissenschaftlichen Ordnungen letztlich widersetzt. Dennoch zeigten sie sich dem Vorgehen nicht abgeneigt, durch Vergleiche zwischen seriellen Quellen grafische Rhizome darzustellen. So verweisen sie in „Milles Plateaux“ auf die Methode Fernand Delignys, der ungeordnete Gesten und Bewegungen verschiedener autistischer Kinder aufgezeichnet und übereinandergeschichtet hat. Deleuze und Guattari sehen in einer solchen Überschichtung von einzelnen Karten die Möglichkeit, darin intersubjektive Verdichtungen zu erkennen (Deleuze/Guattari 1992: 26). Überträgt man dies zurück auf die Diskursanalyse so könnte man sagen, dass „Wucherungen“, die sich in einer solchen Kartografierung um gewisse Begriffe und Themen bündeln, als Diskurse bezeichnet werden können.²⁰ Das „Wuchern der Diskurse“ zeigt sich in der Verdichtung der Verästelungen und Verknüpfungen, die um einen einzelnen Begriff herum entstehen. Solche Übereinanderschichtungen zur Ermittlung von Ballungen, Zentren im Rhizom lassen sich heute korpusgeleitet ermitteln und visualisieren.

4.2 Datengeleitete Verfahren und Visualisierung

Visualisierungen sind dann Mittel der Erkenntnisgewinnung, wenn andere Formen der Repräsentation wie Listen, Tabellen oder Texte zu umfangreich oder zu komplex sind, um als Ganze erfasst und interpretiert werden zu können. Dies ist bei datengeleiteten Analysen der Normalfall. Die Entwicklung von Methoden zur Visualisierung ist daher ein integraler Bestandteil des Forschungsprozesses.

Beispielgebend sind die technischen Disziplinen, wo vor allem in Mathematik, Informatik, Naturwissenschaften und *life-sciences* bei der Analyse komplex vernetzter Daten mit Visualisierungen gearbeitet wird, die in der Medizin den Namen „bildgebende Verfahren“ tragen. Diese haben nicht nur illustrativen Charakter, sondern sind selbst Mittel der Erkenntnisgewinnung, insofern sie die Daten filtern, gewichten und in eine Form bringen, die vom Menschen erfassbar und interpretierbar ist.

Dies lässt sich an einem der ältesten bildgebenden Verfahren in der Medizin – dem Röntgenbild – erläutern. Röntgenstrahlen werden, je nach der Dichte des Gewebes, auf das sie gerichtet werden, teilweise gefiltert. Die Menge der Strahlen, die nach dem Durchgang durch ein Gewebe noch messbar sind, lassen also Rückschlüsse auf dessen

²⁰ Nicht zuletzt hat Deleuze Foucault in seinem ihm gewidmeten Buch als „Kartographen“ bezeichnet, der die Verknüpfungen zwischen der diskursiven und nicht-diskursiven Ebene der Machtausübung skizziert. (Deleuze 1992: 54).

Beschaffenheit zu. Würde die Verteilung der durchgedrungenen Strahlen auf einer Fläche in Form einer Zahlenmatrix vorliegen, wäre diese Matrix angesichts der unzählbaren Menge an Strahlen nicht erfassbar. Durch die Lenkung dieser Strahlen auf einen geeigneten Film (oder ein digitales Detektorensystem) jedoch kann die Verteilung durchgekommener und gefilterter Strahlen visualisiert werden. Es entsteht ein zweidimensionales Bild, das die Durchdringbarkeit (dunkle Bereiche) bzw. Undurchdringbarkeit (helle Bereiche) des Gewebes für Röntgenstrahlen repräsentiert und damit Rückschlüsse auf die Beschaffenheit des Gewebes zulässt. Noch erheblich komplexer wird die Analyse im Fall der Computertomographie: sie kombiniert die Daten vieler Röntgenbilder aus unterschiedlichen Perspektiven zu einem dreidimensionalen Bild des Gewebes.

Visualisierungen sind freilich keine Abbildungen der Wirklichkeit, sondern aufgrund von Relevanzkriterien geordnete Reduktionen von Daten, die auf der Basis gestalterischer Vorgaben visualisiert werden. Beim klassischen Röntgenbild etwa wird ein dreidimensionaler Raum auf eine Fläche reduziert; und in einer Computertomographie werden Repräsentationen von Gewebeteilen mit dem Ziel der Unterscheidbarkeit eingefärbt und nicht ihrer tatsächlichen Färbung entsprechend.

Auch im Bereich der Sozialwissenschaften fassen Visualisierungen über den Bereich der Illustration statistischer Informationen hinaus mehr und mehr Fuß, etwa im Bereich der Netzwerkanalyse. In den Kultur- und Geisteswissenschaften steckt die Entwicklung von durch Visualisierung geleiteten Methoden dagegen noch in den Kinderschuhen. Und dies, obwohl seit der Begründung der digital humanities und der Computerphilologie große digitale Textarchive sowie kultur- und kunstwissenschaftliche Datenbanken entstanden sind.

4.3 Kollokationen und Kollokationsgraphen

Der Begriff der Kollokation ist in der Linguistik durchaus umstritten (vgl. Evert 2009: 1212f.). Einig sind sich jedoch die meisten Definitionen darin, dass es sich bei Kollokationen um rekurrente Kookkurenzen von mindestens zwei lexikalischen Einheiten in einer begrenzten Menge sprachlicher Daten handelt. Typische Beispiele für Kollokationen sind „Weg“ und „einschlagen“ oder „Geschichte“ und „erzählen“, die in sprachlichen Daten jeweils überzufällig häufig miteinander auftreten. Je nach Erkenntnisinteresse und Anwendungsgebiet der Kollokationsanalysen werden weitere Bestimmungen als konstitutive Bestandteile des Kollokationsbegriffs formuliert: etwa im Hinblick auf die Art der Relation zwischen den Wörtern oder im Hinblick auf die Menge der Sprachdaten, in denen die lexikalischen Einheiten gemeinsam vorkommen. So definiert (Bartsch 2004: 76), Kollokationen seien „lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in a direct syntactic relation with each other.“ Evert (2009: 1213f.) dagegen, dem wir in unserem Projekt folgen, unterscheidet einen rein statistischen Kollokationsbegriff von dem Begriff der *multiwordexpression*. Ersterer bezieht sich auf ein in natürlichen Sprachen direkt beobachtbares gemeinsames Auftreten von Wörtern, letzterer basiert auf theoretischen Annahmen über die Semantizität und Pragmatizität der Wortverbindungen und ist somit in höherem Maße von theoriegeleiteten Interpretationen abhängig.

Die Grundlage für die Berechnung von Kollokationen ist die Bestimmung von Kollokatoren. Kollokatoren sind Wörter, die gemeinsam mit einem bestimmten Wort, er

„Basis“, auftreten. Je nach linguistischem Erkenntnisinteresse legt man ein Fenster von n Wörtern links und m Wörtern rechts der Basis fest und wertet dieses Fenster bei jedem Auftreten der Basis im Korpus aus. So erhält man eine Liste von Wörtern, die gemeinsam mit der gesuchten Basis auftreten, sowie ihrer Frequenzen. Die Liste der Kollokatoren alleine ist jedoch noch nicht aussagekräftig. Funktionswörter wie Artikel oder Präpositionen treten als Angehörige von geschlossenen Wortklassen insgesamt sehr viel häufiger auf als Adjektive, Verben oder Nomen. Das häufige gemeinsame Auftreten des Wortes „der“ mit dem Wort „Weg“ reicht nicht dafür aus, „der“ als Kollokation von „Weg“ zu klassifizieren. Um zu überprüfen, ob eine Basis und ein Kollokator überzufällig häufig miteinander auftreten, muss untersucht werden, ob der Kollokator in anderen Kontexten nicht ähnlich häufig auftritt wie im Kontext der Basis. Hierfür teilt man das Korpus in zwei Teilkorpora. Das erste Teilkorpus besteht aus allen Basen und den dazugehörigen Textfenstern, also von den n Wörtern links und m Wörtern rechts der Basis. Das zweite Teilkorpus besteht aus allen restlichen Wörtern. Nun zählt man die Frequenzen des Kollokatorwortes im zweiten Korpus und berechnet, wie häufig es in beiden Korpora relativ zur Wortzahl des jeweiligen Korpus vorkommt. Zur Beantwortung der Frage, ob das gemeinsame Auftreten von Kollokator und Basis als zufällig oder überzufällig gelten kann, setzt man die relativen Frequenzen mit einander in Beziehung: als statistische Verfahren werden unterschiedliche Signifikanztests verwendet.²¹ Je nach Wahl der Parameter (Korpus, Fenstergröße, Grenzwerte, Signifikanztest) erhält man teilweise unterschiedliche Ergebnisse.

Da es in unserem Projekt darum geht, Frames zu identifizieren, die in Zeitungstexten signifikant häufig zusammen vorkommen, ist das Konstrukt der Kollokation, verstanden als ein überzufälliges gemeinsames Auftreten zweier Phänomene, ein geeigneter Anknüpfungspunkt. Jedoch unterscheidet sich die lemma- oder wortformspezifische Kollokationenberechnung von der Berechnung von textspezifischen Framekollokationen in einem wichtigen Punkt: Für erstere ist die Linearität der sprachlichen Daten, d.h. ihre Abfolge, relevant, was sich auch an der Definition einer Wortspanne links und rechts der Basis zeigt; für letztere hingegen sind die Frames, die in einem Text vorkommen, eine ungeordnete Menge. In dieser Menge kann jeder Frame genau einmal vertreten sein, ganz gleich wie umfangreich der Text ist. Bei der lemma- oder wortformspezifischen Kollokationenberechnung ist es hingegen möglich, dass ein Kollokator mehrmals im gleichen Fenster auftritt und die Wahrscheinlichkeit dafür wächst, je größer das Fenster gewählt wird.

Ausgehend von diesen Unterschieden in der Datenstruktur waren bei der Anwendung eines kollokationsanalytischen Verfahrens auf unsere Fragestellung folgende Anpassungen nötig: (1) die Größe des Fensters zur Bestimmung von Kollokatoren liegt nicht fest, sondern entspricht jeweils der Menge der im jeweiligen Text vorkommenden Frames. (2) Als Basis für die Berechnung der Typizität des gemeinsamen Auftretens von Frames wurde nicht die Anzahl der Token in den beiden Teilkorpora gewählt, sondern die Anzahl der Texte mit bzw. ohne gemeinsames Auftreten.²²

Während sich Informationen über die Kollokationen zu einer einzigen Basis leicht als Liste oder Tabelle überblicken lassen, ist die Analyse von Kollokationsnetzen, also Informationen über die Kollokationen mehrerer Basen, die selbst Kollokationen anderer

²¹ Für eine weiterführende Darstellung in Frage kommender statistischen Verfahren vgl. Evert 2009: 1224-1236.

²² Frame- und Kollokationsanalyse wurde mittels selbst geschriebener Software durchgeführt, ebenso wie die Überführung der Einzelergebnisse in ein Kollokationsnetz. Das Setzen des Graphen erfolgte mit Hilfe des GraphViz-Software-Pakets (Ellson et al. 2003).

Basen sein können, bei einer nennenswerten Menge an Verbindungen ohne Visualisierung nicht mehr möglich. Basis-Kollokationen-Verbindungen und Kollokationsnetze lassen sich aber beispielsweise als gerichtete Graphen visualisieren. Graphen sind – einfach gesprochen – eine Menge von Punkten („Knoten“ / „vertices“ / „nodes“), zwischen denen Linien („Kanten“ / „edges“) verlaufen. In unserem Projekt repräsentieren die Punkte Frames (bei den späteren Untersuchungen zum Familien-Frame auch Lemmata), die Kanten signifikante Kookkurrenzen zwischen zwei Knoten. Es gibt freilich unzählige Möglichkeiten einen Graphen mit mehreren Knoten und Kanten zu zeichnen.²³ Insbesondere für komplexe Netze haben Informatiker und Mathematiker verschiedene Algorithmen entwickelt, die automatisiert das Layout eines Graphen erzeugen. Dabei kommen auch ästhetische Kriterien zur Anwendung, damit die Lesbarkeit des Graphen gesichert und optimiert wird. Dies sind unter anderen:

1. die Minimierung der Überschneidungen von Kanten,
2. die Minimierung der Fläche,
3. die Minimierung der Länge der Kanten, (vgl. Fleischer / Hirsch 2001: 19f.)
4. die Verhinderung des Überschreibens von Knoten in sehr großen Graphen.

Da über die Struktur der Graphen in unserem Projekt nichts vorgängig bekannt ist, wir beispielsweise keine Hierarchien oder Verwandtschaften voraussetzen können, bietet sich die Verwendung einer Layout-Methode an, die prinzipiell für alle Arten von Graphen geeignet ist. Die Wahl fiel auf die forced-based-Methode (Kräftediagramm), die einen Graphen als ein physikalisches System modelliert, das von Kräften (forces) zwischen den Knoten gebildet wird (vgl. Fleischer / Hirsch 2001: 20). Man kann sich dabei die Kanten als Sprungfedern und die Knoten als Ringe vorstellen, in die die Sprungfedern eingehakt sind. Erlaubt man einem solchen physischen Graphen, sich ohne Beschränkung zu bewegen, wird er in einen Zustand kommen, in dem die potentielle Energie in den Federn möglichst gering ist (vgl. Landgraf 2001: 173). Die für die Optimierung des Graphen-Layout hilfreiche Analogie dabei ist, dass ein Graph dann Ansprüche an seine Informativitäts- und die oben genannten Lesbarkeits-Kriterien am besten erfüllt, wenn er sich in einem Zustand der Entspannung, d.h. in einem Zustand mit geringer energetischer Ladung befindet (vgl. Brandes 2001: 71).

Die Implementierung von Algorithmen zur Sicherung der Lesbarkeit der Graphen hat zur Folge, dass nicht alle Aspekte eines Graphen als Visualisierung relevanter Informationen gelesen werden können. So sind beispielsweise die Kantenlängen und die Winkel zwischen zwei Kanten nicht bedeutungstragend. Der Algorithmus sorgt jedoch dafür, dass zwei oder mehr Knoten, die mit einer gleichen oder sehr ähnlichen Menge anderer Knoten verbunden sind, in relativer Nähe zueinander dargestellt werden. Verdichtungsbereiche im Graphen lassen also auf eine Gruppe von Knoten mit ähnlichen Vernetzungsmerkmalen schließen.

Kollokationsgraphen bieten damit die Möglichkeit, Frames mit ähnlichen Vernetzungsmerkmalen anhand von Verdichtungsräumen im Graphen visuell zu identifizieren. Zudem kann es interessant sein, jene Kanten zu untersuchen, durch die verschiedene Frame-Cluster miteinander verknüpft sind. Aufgrund der Komplexität der Daten wären diese Analysen ohne Visualisierung kaum möglich.²⁴

²³ Eine Übersicht bieten die Standardwerke und Kaufmann / Wagner 2001 und Tamassia in Druck.

²⁴ Weil wir teilweise Graphen mit mehr als 5.000 Kanten setzen mussten, haben wir auf einen Force-directed Algorithmus zurückgegriffen, der sehr große Graphen in hoher Geschwindigkeit layoutet und über einen speziellen Algorithmus zur Verhinderung des Überschreibens von Knoten verfügt: den Scalable Force Directed Placement-Algorithmus (vgl. Hu 2005).

Im Folgenden wollen wir zeigen, wie Kollokationsgraphen genutzt werden können, um die Qualität von Umbrüchen zu untersuchen.

5. Datengeleitete Frameanalyse am Beispiel des ZEIT-Archivs II: Wandel von Framekollokationen in Umbruchzeiten

Ausgehend von den in Abschnitt 3 berechneten Umbruchjahren haben wir berechnet, welche Framekonstellationen sich im Zuge der Umbrüche besonders stark verändert haben. Im Folgenden sollen zunächst die Analyseergebnisse für den Umbruch in den späten 2000er Jahren anhand eines Vergleichs der Framekollokationen in den Jahren 2004-2007 und 2008-2011 vorgestellt werden, ehe der Umbruch zwischen den Jahre 1991 und 1992 genauer untersucht wird.

5.1 Die Wiederkehr des Politischen: Die Umbruchjahre 2008-2011

Um die Qualität des Umbruchs, der sich in den Jahren 2008 bis 2011 vollzogen hat, genauer zu untersuchen, haben wir für die Umbruchjahre und für einen gleich langen Zeitraum vorher, die Jahre 2004 bis 2007, je einen Kollokationsgraphen berechnet. Dabei haben wir sämtliche Texte aus den jeweiligen Untersuchungszeiträumen einer automatisierten Frameanalyse unterzogen und daran anschließend untersucht, welche Frames signifikant häufig im jeweiligen Zeitraum mit einander in Einzeltexten kookkurrieren. Um die Unterschiede der beiden Zeitabschnitte leichter vergleichen zu können, haben wir Knoten und Kanten beider Rhizome zu einem gemeinsamen Frame-Kollokationsgraphen vereinigt, wobei die für die Umbruchjahre spezifischen Kanten hervorgehoben wurden (Abbildung 10).

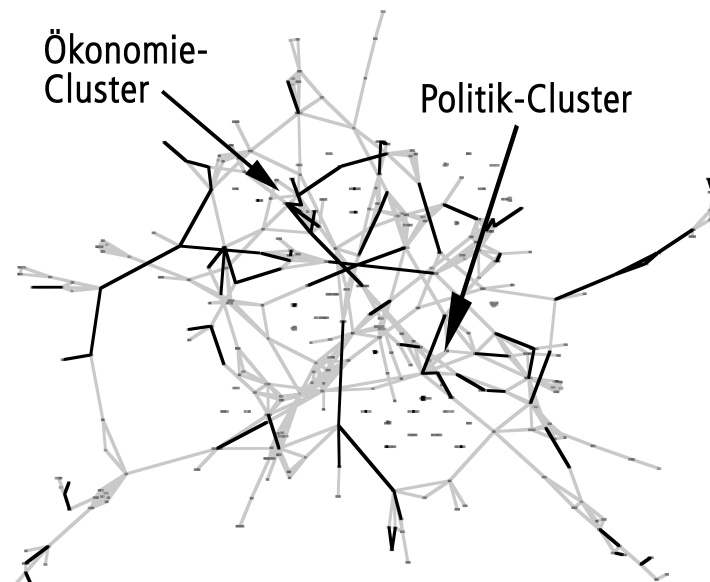


Abbildung 10: Frame-Kollokationen im ZEIT-Archiv der Jahre 2004 bis 2011. Spezifische Frame-Kollokationen der Jahre 2008 bis 2011 sind schwarz hervorgehoben.

Die visuelle Analyse zeigt in mindestens zwei Clustern einschneidende Veränderungen, die sich bei näherer Betrachtung der in ihnen verdichteten Frames als Politik- und Ökonomie-Cluster bezeichnen lässt.

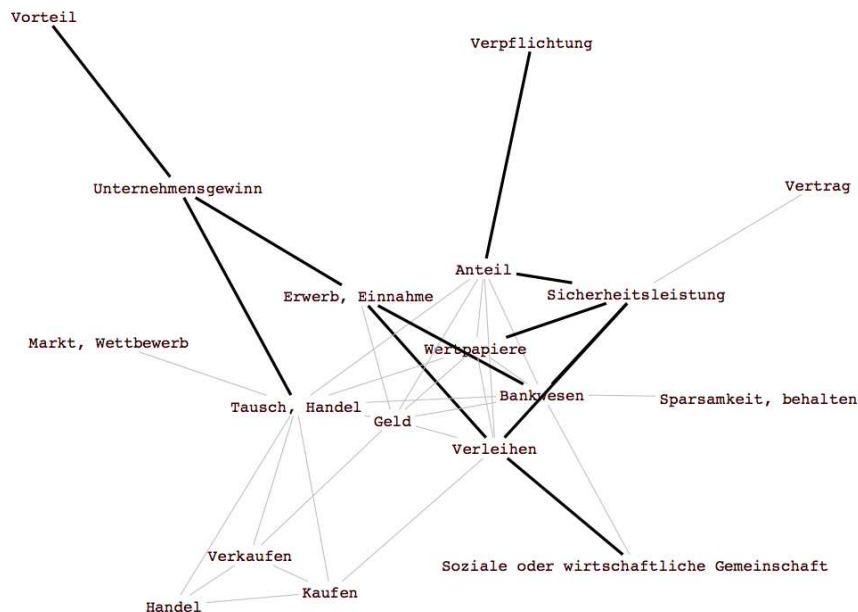


Abbildung 11: Ökonomie-Cluster im Frame-Rhizom (ZEIT-Archiv der Jahre 2004 bis 2011). Spezifische Frame-Kollokationen der Jahre 2008 bis 2011 sind schwarz hervorgehoben.

Im Cluster „Ökonomie“ lassen sich deutliche Verschiebungen feststellen: So geraten die Frames „Verkaufen“, „Handel“, „Kaufen“ und sogar „Markt, Wettbewerb“ in den Hintergrund gegenüber „Verleihen“, „Sicherheitsleistung“, „Verpflichtung“ und „Unternehmensgewinn“, der mit „Tausch“ und „Handel“ assoziiert ist. Stärker gewichtet wird auch die „soziale oder wirtschaftliche Gemeinschaft“, welche über „Verleihen“ an das Cluster angebunden ist.

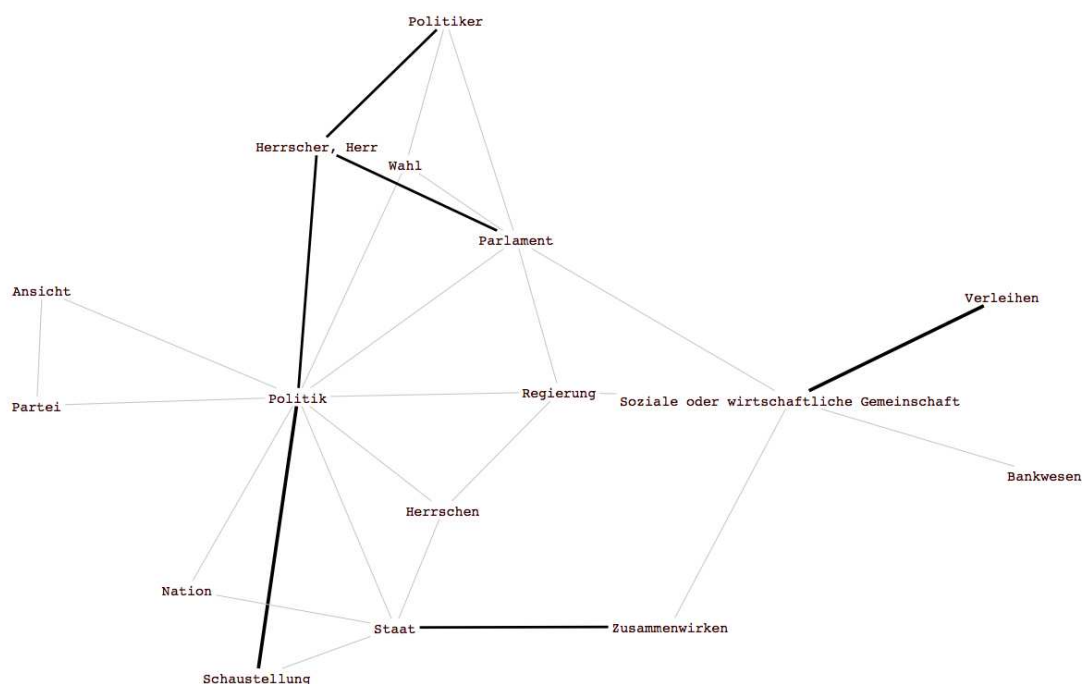


Abbildung 12: Politik-Cluster im Frame-Rhizom (ZEIT-Archiv der Jahre 2004 bis 2011). Spezifische Frame-Kollokationen der Jahre 2008 bis 2011 sind schwarz hervorgehoben.

Im Cluster „Politik“ wird zugleich sichtbar, dass im Vergleich zwischen 2008-2011 und 2004-2007 nicht nur die Verbindung zu „verleihen“ typischer wird, sondern auch das „Zusammenwirken“ zwischen „Staat“ und „Wirtschaft“. Es lässt sich ein krisenbedingtes, vorübergehendes Erstarken der Politik z.B. auch in stärkeren, typischen Verbindungen zwischen den Frames „Parlament“ und „Herrschaft“ erkennen. Ähnliche Tendenzen lassen sich im nicht abgebildeten Gesamtrhizom auch zwischen „Gesetz“ und „Stärke“ feststellen. Zudem sieht man eine weitere Veränderung im Gesamtrhizom, wo die starke Verbindung zwischen „Freiheit“ und „Pflicht“, die 2004 aktiv ist, abgelöst wird durch die Verbindung „Freiheit“ und „Absolut“, welches wiederum verbunden ist mit „Sorgfalt“ und einer zunehmenden Nähe zu „Gesetz“ und auch „Pflichtverletzung“.

Es ist anzunehmen, dass sich darin die Tendenz spiegelt, nach der Wirtschaftskrise mehr Einbindung wirtschaftlicher Freiheit in klare Regelwerke zu fordern. Auch in den zuvor erwähnten Verbindungen zeigt sich eine zumindest für die Jahre von 2008 bis 2011 anhaltende Konjunktur politischer Einbettung wirtschaftlicher Prozesse an. Im Rhizom erscheint die Wirtschaft gegenüber dem Staat in einer Bringschuld, was Frames wie „Verpflichtung“ und „Sicherheitsleistung“ deutlich machen.

5.2 Von der Staatsnation zur Kulturnation: Das Umbruchjahr 1992

Auch für die Jahre nach der Deutschen Einheit haben wir je einen Frame-Kollokationsgraphen für den ersten Zeitabschnitt (1991) und einen für den zweiten Zeitabschnitt (1992) berechnet und die beiden Rhizome zu einem gemeinsamen Graphen vereint, in dem die spezifischen Frame-Kollokationen der Umbruchjahre hervorgehoben sind. Wie Abbildung 13 zeigt, lassen sich in diesem Graphen drei Cluster identifizieren, in denen besonders viele für das Umbruchjahr 1992 spezifische Frame-Kollokationen verdichten.

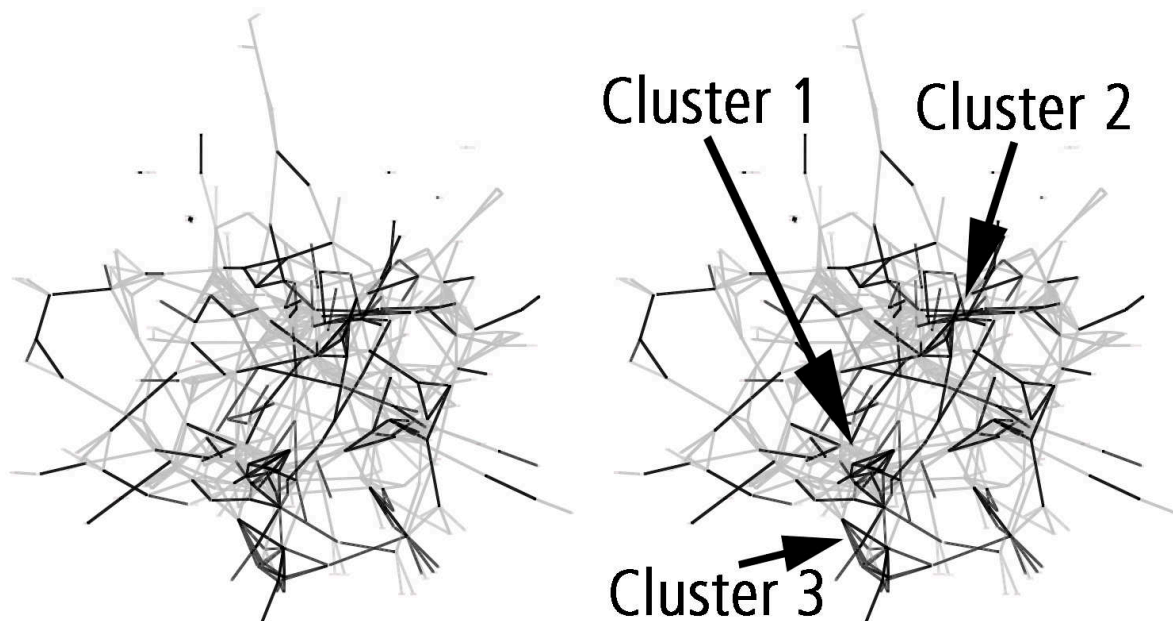


Abbildung 13: Frame-Kollokationen im ZEIT-Archiv der Jahre 1991 und 1992 ohne Beschriftung (links) und mit Beschriftung (rechts). Spezifische Frame-Kollokationen des Jahrs 1992 sind schwarz hervorgehoben.

Besonders interessant erscheint uns das Cluster 2, das sich um die Frames „Freiheit“ und „Nation“ formiert.

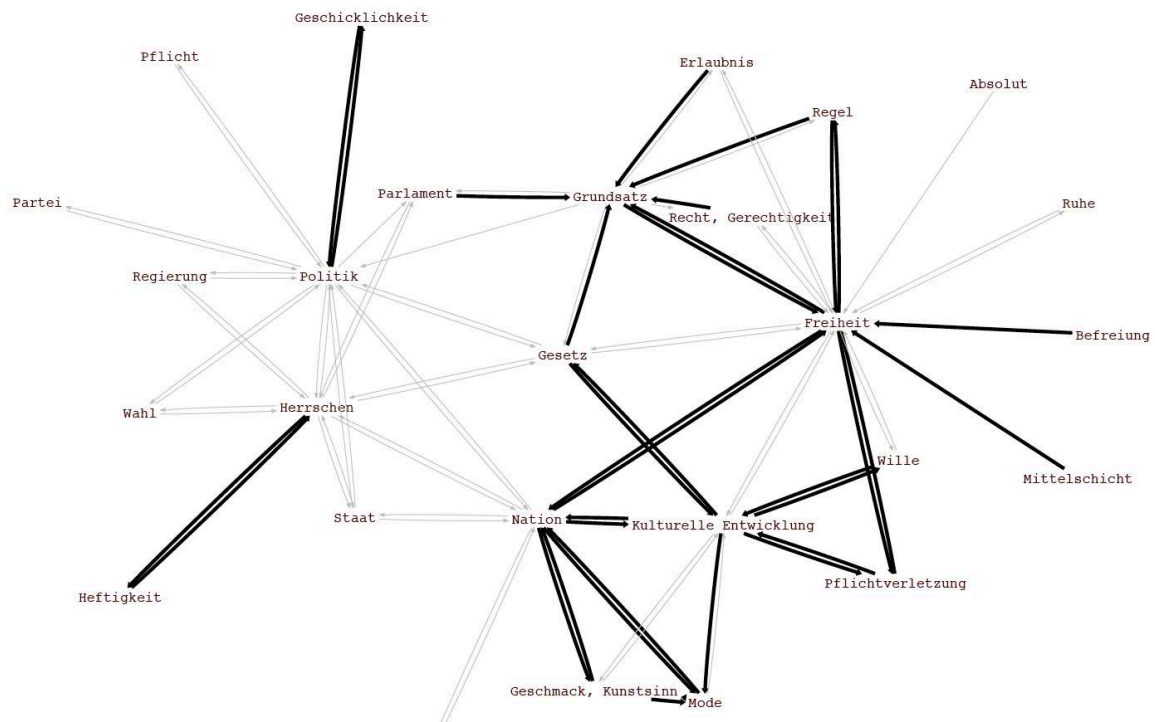


Abbildung 14: Frame-Kollokationen im ZEIT-Archiv der Jahre 1991 und 1992, Cluster 2. Spezifische Frame-Kollokationen des Jahrs 1992 sind schwarz hervorgehoben.

Der Frame „Nation“ ist dabei erwartbar stark verbunden mit dem „Freiheits“-Frame, welcher wiederum soziologische Frames wie „Mittelschicht“ und politisch-rechtliche wie „Grundsat“ um sich bündelt aber auch jenen der „Befreiung“. Zugleich entsteht 1992 um den Frame „Nation“ eine Verbindung mit Frames wie „Mode“ und „Geschmack“, „Kunstsin“, „Kulturelle Entwicklung“. Damit öffnen sich die Verbindungen die der Frame „Nation“ eingeht im Gegensatz zur Situation im Jahr 1991: Im jahresspezifischen Rhizom finden sich keine solchen typischen Bezüge: „Nation“ verbindet sich mit „Herrschen“ und „Politik“. Darin zeigt sich ein Wandel der Konstruktion des Nationalen von einer auf politischem Handeln gründenden staatlichen Einheit (1991) hin zu einer stärker über kulturelle Werte definierten nationalen Gemeinschaft (1992). Zugleich macht das Rhizom Erfahrungsmöglichkeit der Nation und ihrer Wiedervereinigung auf der Ebene persönlicher sinnlicher Konsumerfahrung sichtbar.

6. Datengeleitete Frameanalyse am Beispiel des ZEIT-Archivs III: Analyse thematischer Frames

Die Frameanalyse eignet sich zwar gut für die Identifizierung von Umbruchzeiten und erlaubt Rückschlüsse auf die Qualität der Veränderungen, sie ist aber insgesamt ein eher grobes Messinstrument. Im Folgenden sollen zwei Methoden aufgezeigt werden, wie Veränderungen einzelner diskursiver Stränge präziser untersucht werden können.

6.1 Der Tugend-Frame 1946-1966 und 2007-2011 im Vergleich

Für die folgende exemplarische Analyse haben wir den Tugend-Frame ausgewählt, weil er von allen Frames, die die Referenz auf Werte repräsentieren, zur Gegenwart hin am meisten an Relevanz gewinnt. Abbildung 15 zeigt auch, dass der Tugend-Frame vom Anfang des Untersuchungszeitraums bis Mitte der 1960er Jahre bereits Konjunktur hatte.

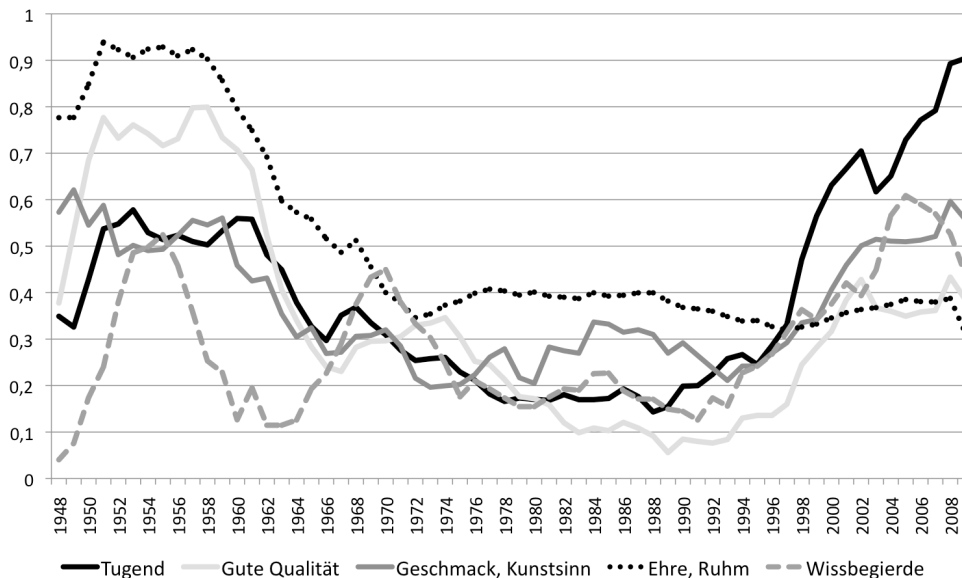


Abbildung 15: Jahresweise Verteilung ausgewählter wertspezifischer Frames im Printarchiv der ZEIT.

Uns interessierte daher, welche Themen mit dem Tugend-Frame in den unterschiedlichen Zeiträumen assoziiert waren. Um herauszufinden, welche Themen mittels des Tugend-Frames verhandelt wurden haben wir alle Texte, die den Tugend-Frame enthalten, zu zeitspezifischen Subkorpora zusammengefasst. Diese Subkorpora wurden dann auf Veränderungen in den Lemma-Kollokationen hin untersucht.

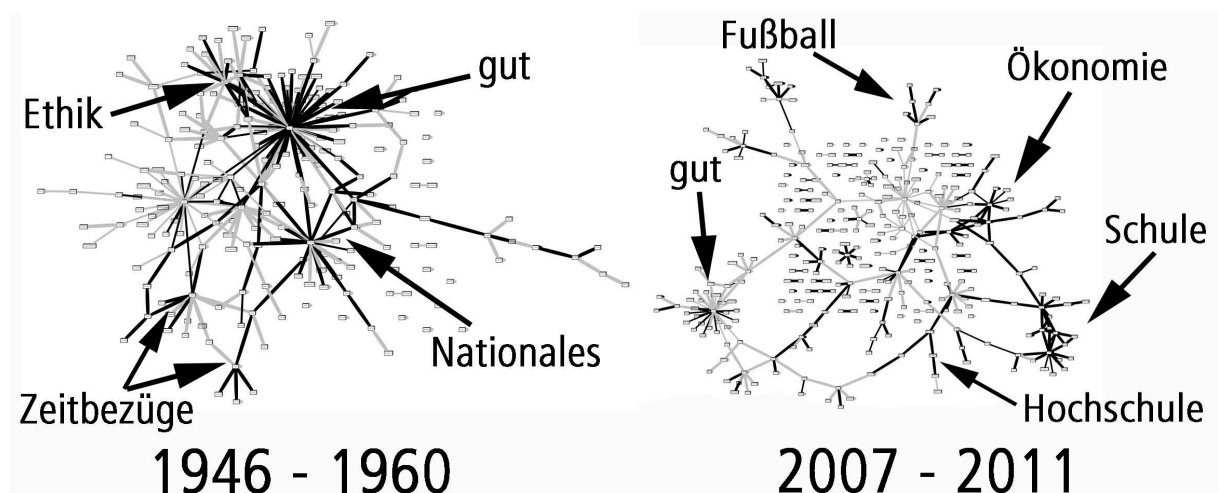


Abbildung 16: Rhizome aus Lemma-Kollokationen aus allen Texten, in denen der Tugend-Frame vorkommt: ZEIT-Archiv 1946-1960 und 2007-2011 im Vergleich.

1946-1960 finden sich die Hauptbündelungen des Rhizoms, neben jenen zu „gut“, auf die ich später eingehen möchte, um „Mensch“, „deutsch“, „geben“, sowie um

zeitabschnittbezogene Wörter wie „Zeit“ und „Jahr“ und „jetzt“. Im Gegensatz dazu bündelt sich das Rhizom von 2007-2011 stark auf bildungsbezogene Lexeme. Die Hauptbündelungen von 2007-2011 unterscheiden sich deutlich zu jenen 60 Jahre zuvor: Ein zentraler Bündelungsbereich zeigt sich um „Schüler“ und „Kind“. Eine kleinere, neue Bündelung im Tugendframe findet sich auch um „Hochschule“, das mit „deutsch“ verbunden ist, welches jedoch keine zentrale Bündelung mehr darstellt. Die Bündelung um „Schüler“ wiederum ist mit „Leistung“ auch mit dem Lexem „Mitarbeiter“ verbunden. Dieser hängt mit einer weiteren Bündelung um „Geld“ zusammen, in dem der Bereich Währungspolitik, der bereits 1945-50 marginal existierte, ausgebaut wurde. Interessant erscheint uns insbesondere die Verschiebung zu sein, dass nicht mehr der „Mensch“ einen der Fokussierungspunkte darstellt, sondern das „Kind“, der zu bildende, unfertige Mensch, der als „human resource“ angebunden an semantische Felder wie Arbeit und Währungspolitik ist.

Die Bündelung „Mensch“ scheint jedoch auch noch im Rhizom von 2007-2011 auf, wenn auch in einer marginaleren, weniger zentralen Position. Ein Blick darauf, wie sich die Kollokationen verändert haben, zeigt, dass es durchaus Konstanten gibt: „Mensch“ zeigt sich auch noch 2007-2011, wie 1946-1950, mit religiösen Lexemen, wie „Gott“ und „Glaube“, aber auch mit „böse“ und „Wille“ verknüpft. Weggefallen allerdings sind „modern“ und „leben“. 2007-2011 hinzugekommen sind allerdings „glücklich“ und „Möglichkeit“. Insbesondere das auftauchen von „Möglichkeit“ in Kollokation mit „Mensch“ passt sich semantisch ein in ein Gesamtrhizom, in dem der junge, unfertige, zu bildende Mensch überdeutlich im Zentrum steht: Der Mensch wird, in der „Multioptionsgesellschaft“ (Gross 1994) und im Zeitalter der „permanenten Weiterbildung“ (Deleuze 1990: 262) als unbestimmtes, formbares betrachtet. Damit einher geht auch der Schwund von auf feste kulturelle Werte bezogenen Lexemen, die sich um „gut“ versammeln. Die typischen Kollokationen von „gut“ mit „Wert“, „Buch“, „Zeitung“, „Musik“, „Geschmack“, „Ton“, „Beispiel“, „künstlerisch“ und „moralisch“ verschwinden. Darin lässt sich unschwer das Verschwinden bildungsbürgerlicher Werthorizonte erkennen. Einige der Lexeme finden sich 2007 an anderer Stelle wieder: Der „Wert“ z.B. scheint nur noch in der Bündelung um Finanz- und Währungsfragen auf, ein Buch wird 2007-2011 eher als „neu“ denn als „gut“ bezeichnet, im Vordergrund steht nicht mehr der Inhalt, sondern das Erscheinen auf dem Markt. Das Verb „fördern“ verschiebt sich ebenfalls weg von „gut“: 2007-2011 wird, wie erwähnt, nicht mehr das „Gute“ gefördert, sondern das „Kind“.

Hingegen zeigen sich auch, wie bei der neu auftauchenden Kollokation von „Mensch“ mit „glücklich“ auch mehrere genussorientierte Lexeme, die mit „gut“ verbunden sind: „Humor“ und „Laune“ kommen zur Kollokationsbündelung um „gut“ hinzu: Die eigene Befindlichkeit wird wichtiger was sich auch in dem Verschwinden von „Geschmack“ anzeigt, das seinen Ersatz in „schmecken“ findet: Nicht mehr der gesellschaftlich definierte „gute Geschmack“ oder „gute Ton“ ist entscheidend, sondern dass dem Konsumenten selbst etwas gut „schmeckt“.

In den angedeuteten Verschiebungen zwischen den Zeiträumen 1946-1960 und 2007-2011 zeigt sich deutlich die Veränderung der in der ZEIT beschriebenen Gesellschaft weg von einem klar definierten Wertebasis hin zu einer leistungsorientierten Bildungsgesellschaft, in der Bildung weniger eine traditionelle Geschmacksorientierung zum Ziel hat, sondern einerseits das Herausbilden und Ausschöpfen von wirtschaftlichen „Möglichkeiten“ und andererseits die Ermöglichung subjektorientierter Genussmaximierung.

6.2 Der Familien-Frame von 1946 bis 2011

Eine andere Möglichkeit, sich dem diskursiven Funktionswandel thematischer Frames anzunähern, ist die Analyse der Kollokationen zum framekonstitutiven Wortschatz in je zeitspezifischen Teilkorpora – ein Verfahren, das wir im Folgenden an der Entwicklung des Familien-Frames illustrieren wollen. Der Familienframe gehört zu jenen Frames, dessen relative Frequenz im Laufe der Jahre stark zunimmt und zwar besonders in den letzten 15 Jahren (vgl. Abbildung 4). Abbildung 17 zeigt die Lemma-Kollokationsgraphen, die sich durch die Berechnung der Kollokationen zu allen Lemmata, die den Familien-Frame konstituieren, in unterschiedlichen Zeitabschnitten ergeben haben.

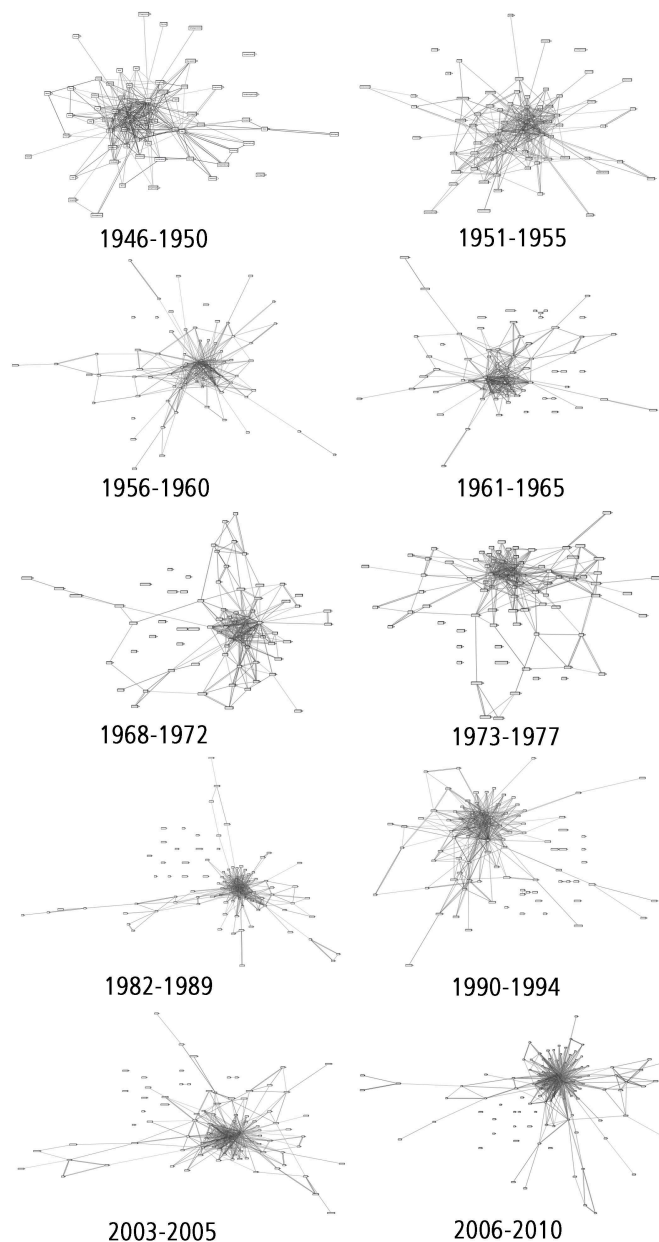


Abbildung 17: Lemma-Kollokationen zum sachbereichsspezifischen Vokabular des Familien-Frames in unterschiedlichen Zeiträumen im ZEIT-Archiv.

Als erster Befund lässt sich feststellen, dass der Kernbereich des Familienframes sich in allen Zeiten gleich bleibt: „Vater“, „Mutter“, „Sohn“, „Tochter“ stehen über alle Jahrzehnte im Zentrum, ganz im Mittelpunkt steht hier immer schon das „Kind“. Weiter permanent dazu gehören „erzählen“ und „sterben“. Nach 1990 zunehmend ausgelagert aus dem Kernbereich des Familienrhizoms erscheinen Wörter wie „Oma“, „Opa“, „Enkel“ und „Grossmutter“. Erstaunlich ist zugleich, wie stark die Familie kontinuierlich mit einem eigenen Haus verbunden wird. Während „Haus“ sich bis 2010 im Kernbereich befindet, taucht die „Wohnung“ nur 1961 überhaupt am Rande des Zentrums im Rhizom auf, um danach bis 2002 wieder zu verschwinden. Im Folgenden soll auf einige sich verändernde Bereiche eingegangen werden, an denen sich auch exemplarisch Aussagen über die Aus- und Rückbildung von Diskursen in Massenmedien festmachen lassen.

Einerseits gibt es Bereiche, die stark zeitabhängigen Konjunkturen der höheren Problematisierung²⁵ ausgesetzt scheinen und später keine Ausdifferenzierung in der Zeitung erkennen lassen und insofern den Sprung zu einer anhaltenden, breiteren gesellschaftlichen Thematisierung nicht geschafft zu haben scheinen: So erscheint im Umfeld von „Erziehung“ nach 1968 „sexuell“, nach 1973 kommt „Sexualerziehung“ dazu, „Sexfilm“ ist verbunden mit „Braut“ und „Ehefrau“. Beide Verknüpfungen bilden keine eigenständigen Bereiche aus und verschwinden nach den 1970er Jahren wieder gänzlich, genauso wie die „Wohngemeinschaft“, welche die Presse nur in dieser kurzen Phase nach 1968 ausgiebig zu beschäftigen scheint. Ähnlich zeitabhängig scheinen die Verknüpfungen zu sein, die der Begriff „Angehörige“ in diesem deutschen Kontext eingeht. So erscheint dieser nach dem Zweiten Weltkrieg bis 1956 verbunden mit „deutsch“, eine Verbindung die erst nach 1990, bzw. nach dem Mauerfall 1989 wieder kurz auftaucht.

Eine Entwicklung von kurzer, intensiver im Diskurs sichtbarer Beobachtung mit nachherigem Rückgang der rhizomatischen Ausbündelung lässt sich um den Begriff „Adoptivkind“ ausmachen, der sich ebenfalls nach 1968 etwas ausdifferenziert und näher zum Kernrhizom des Familienframes heranrückt. Im Verlauf der Jahre lässt sich auch erkennen, was das prototypische Adoptivkind ausmachte, war es nach 1968 das „Heimkind“, das adoptiert wurde, so ist es nach 1990 eher das „Scheidungskind“, das neben den leiblichen Kindern adoptiert wird. Nach 2006 steht „Adoptivkind“ wieder ohne weitere Verknüpfungen da, scheint also zum möglichen Bestand der Familie zu gehören, ohne aber weiter thematisiert zu werden zu müssen.

Ein besonders interessanter Fall ist die Entwicklung des Diskursbereichs der Scheidung. Um den Begriff „Ehe“ besteht bereits 1946 ein eigener Bündelungsbereich, die Sphären der Familie und der Ehe erscheinen zwar als stark verbunden, aber dennoch als eigenständig. Auf Grund der zunehmenden Problematik der Ehescheidung bildet sich um „Ehe“ und „Scheidung“ herum ein bis in die 1990er Jahre wachsender Diskurs. 1946-1950 stehen nur die Wörter „Scheidung“, „geschieden“ und „Ehescheidung“ stark marginal am Rande des Rhizoms. 1951 kommt neben „Ehefrau“ „Scheidungsklage“ hinzu. 1956-1960 erscheint zum ersten Mal „Gütertrennung“ neben den bisherigen Begriffen. Der Bereich verrechtlicht zunehmend, es zeigen sich neu „Eherecht“, „Ehegesetz“, „Scheidungsurteil“ und „Ehebruch“. Nach 1968 kommen „Strafrechtsreformgesetz“, „Härteklause“, „Zertrüttungsprinzip“, „Zerrüttung“ hinzu, die auf aktuelle juristische Reformen hinweisen. Neben dieser erkennbaren rechtlichen Ausdifferenzierung des Scheidungsbereichs lassen sich auch Anzeichen einer zunehmenden psychologischen Problematisierung erkennen. Diese verweist auf die

²⁵ Zum Begriff der „Problematisierung“ vgl. Foucault 1985: 158.

Herausbildung eines weiteren Berufsbereichs, der sich jenseits des Juristischen professionell mit der Scheidungsproblematik auseinandersetzt. Während die juristische Ausdifferenzierung des Diskurses nach den 1990er Jahren eher abnimmt, erscheinen hier erstmals Begriffe wie „Mediation“, „Trialog“, aber immer auch noch „Notar“, „Familiengericht“. Nach 1998 scheint der Bereich um „Scheidung“ eher zurückzugehen. Der diskursive Problematisierungsbedarf in den Massenmedien geht zurück, der Bereich geht mehr oder weniger zurück auf Grundbegriffe wie „Elternteil“, „Scheidungskind“ und „Unterhalt“.

Gerade an diesem sich Auswachsen von diskursiven Problematisierungsbereichen, das gefolgt ist von einem Rückzug, kann exemplarisch gezeigt werden, was die Rhizomanalyse leisten kann: Einerseits kann sie die Aufmerksamkeit für einzelne Bereiche wecken, sie vermag aber auch Konjunkturen der Problematisierung zu erfassen. Wie der Rückgang zu bewerten ist, lässt sich vermuten: So ist der Rückzug der Scheidungsproblematik in den Texten der „Zeit“ kaum ein Anzeichen dafür, dass sich Scheidungen vermindert haben, sondern vielmehr dafür, dass sie nicht mehr als auffälliges, neues gesellschaftliches Problem betrachtet werden und allfällige Diskussionen in der psychologischen und juristischen Fachpresse weitergeführt werden.

7. Fazit

Die digitale Revolution beschert der Sprachwissenschaft ein immer größeres Reservoir an Daten, wachsende Rechnerleistungen und Fortschritte in Informatik, Computerlinguistik und digitaler Sprachverarbeitung schaffen immer bessere Möglichkeiten, diese Daten effizient zu verarbeiten und präziser zu analysieren. Datengeleitete Zugänge eröffnen völlig neue Perspektiven auf unseren Forschungsgegenstand und haben das Potenzial, alternative Kategorien seiner Beschreibung und Erklärung hervorzubringen. Der vorliegende Text ist ein Plädoyer dafür, von diesen Möglichkeiten auch in der linguistischen Diskursanalyse verstärkt Gebrauch zu machen. Wir sind überzeugt, dass an Arbeiten auf dem Gebiet der Gesellschaftsanalyse qua Sprachanalyse künftig immer mehr die Forderung nach einer breiten empirischen Basis („big data“) herangetragen werden wird, einer empirischen Basis, die sich einer qualitativen Bearbeitung von vornherein verschließt. Diese Forderung ergibt sich aber nicht alleine aus der Verfügbarkeit der Daten, sondern aus dem gerechtfertigten Wunsch nach einer möglichst großen Reichweite und Verallgemeinerbarkeit der Ergebnisse.

Die im vorliegenden Aufsatz dargestellten Methoden verstehen wir als erste Schritte auf einem offenen wissenschaftlichen Feld und nicht als Baustein einer Methodologie. Wir hoffen, dass viele weitere Wissenschaftlerinnen und Wissenschaftler dieses Feld betreten und mit ihren Ideen fruchtbar machen. Sie müssen die Bereitschaft mitbringen, sich auf technische Fragestellungen einzulassen. Im Idealfall lernen sie programmieren, denn der autonome Umgang mit Technik ist die Grundbedingung für eine selbstbestimmte Forschungstätigkeit im digitalen Zeitalter. Wir sind optimistisch, dass in diesem Fall die Geschichte der linguistischen Diskursanalyse ein Erfolgsgeschichte bleiben wird.

Literatur

- Anderson, Chris (2008): The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. In: Wired Magazine 16.07. Online: In: Warnke, Ingo (Hg.): Diskurslinguistik nach Foucault. Berlin/New York: de Gruyter, S. 53-80 (Linguistik – Impulse & Tendenzen; 25).
Online: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory/
- Angermüller, Johannes (2007): Diskurs als Aussage und Äußerung – die enunziative Dimension in den Diskurstheorien Michel Foucaults und Jacques Lacans. In: Warnke, Ingo (Hg.): Diskurslinguistik nach Foucault. Berlin/New York: de Gruyter, S. 53-80 (= Linguistik – Impulse & Tendenzen 25).
- Bartsch, Sabine (2004), Structural and Functional Properties of Collocations in English. Tübingen: Narr.
- Böke, Karin / Jung, Matthias / Niehr, Thomas / Wengeler, Martin (2000): „Vergleichende Diskurslinguistik. Überlegungen zur Analyse internationaler und intralingualer Textkorpora“, in: Niehr, Thomas / Böke, Karin (Hg.): Einwanderungsdiskurse. Vergleichende diskurslinguistische Studien. Wiesbaden: Westdeutscher Verlag, S. 11–36.
- Brandes, Ulrik (2001): Drawing on Physical Analogies. In: Michael Kaufmann / Dorothea Wagner (Eds.): Drawing Graphs. Methods and Models. Berlin / Heidelberg / New York et al.: Springer. S. 71-86.
- Belica, Cyril (1996): Analysis of temporal changes in corpora. In: International journal of corpus linguistics 1/1, S. 61–73.
- Belica, Cyril (2001): Kookkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs.
- Bubenhofer, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. (Sprache und Wissen). Berlin, New York.
- Bubenhofer, Noah / Scharloth, Joachim (in Druck): Korpuslinguistische Diskursanalyse: Der Nutzen empirisch-quantitativer Verfahren. In: Warnke, Ingo H./Meinhof, Ulrike/Reisigl, Martin (Hrsg.): Diskurslinguistik im Spannungsfeld von Deskription und Kritik. Berlin/New York: de Gruyter.
- Bublitz, Hannelore et al. (Hrsg) (1999): Das Wuchern der Diskurse: Perspektiven der Diskursanalyse Foucaults. Frankfurt am Main: Campus.
- Busse, Dietrich (1987): Historische Semantik. Analyse eines Programms. Stuttgart: Klett-Cotta.
- Busse, Dietrich (2006): Text – Sprache – Wissen. Perspektiven einer linguistischen Epistemologie als Beitrag zur Historischen Semantik. In: Scientia Poetica 10, S. 101–137.
- Busse, Dietrich / Teubert, Wolfgang (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: Busse, Dietrich / Hermanns, Fritz / Teubert, Wolfgang (Hrsg.): Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik. Opladen: Westdeutscher Verlag. S. 10-28.
- Dahinden, Urs (2006): Framing. Eine integrative Theorie der Massenkommunikation. Konstanz: UVK-Verlagsgesellschaft.
- Dornseiff (2004): Der deutsche Wortschatz nach Sachgruppen. Berlin, New York: Walter de Gruyter.

- Deleuze, Gilles/Parnet, Claire (1980): Dialoge. Frankfurt am Main: Suhrkamp.
- Deleuze, Gilles (1990): Postskriptum über die Kontrollgesellschaften, in: Deleuze, Gilles: Unterhandlungen 1972-1990. Frankfurt am Main: Suhrkamp. S. 254-262.
- Deleuze, Gilles (1992): Foucault. Frankfurt am Main: Suhrkamp.
- Deleuze, Gilles/Guattari, Félix: Tausend Plateaus. Berlin: Merve 1992.
- Deleuze, Gilles/Guattari, Felix (1996): Was ist Philosophie? Frankfurt am Main: Suhrkamp.
- Ellson, J. / Gansner, E. R. / Koutsofios, E. / North, S. C. / Woodhull, G. (2003): Graphviz and Dynagraph – Static and Dynamic Graph Drawing Tools, in: M. Junger, M. / Mutzel, P. (Hrsg.): Graph Drawing Software. S. 127-148.
- Evert, Stefan (2009): Corpora and Collocations. In: Lüdeling, Anke / Kytö, Merja (Hrsg.): Corpus Linguistics: An International Handbook. Vol. 2. S. 1212-1248. (= Handbooks of linguistics and communication science; 29.2)
- Fleischer, Rudolf / Hirsch, Colin (2001): Graph Drawing and Its Applications. In: Michael Kaufmann / Dorothea Wagner (Hrsg.): Drawing Graphs. Methods and Models. Berlin / Heidelberg / New York et al.: Springer. S. 1-22.
- Foucault, Michel (1985): Geschichte der Sexualität. In: Ästhetik und Kommunikation 57/58, S. 157-164.
- Goffman, Erving (1974): Frame analysis: An essay on the organization of experience. New York: Harper and Row.
- Gross, Peter (1994): Die Multioptionsgesellschaft. Frankfurt am Main: Suhrkamp.
- Hu, Yi Fan (2005): Efficient and high quality force-directed graphdrawing. The Mathematica Journal, 10, S. 37-71.
- Kämper, Heidrun (2007): Linguistik als Kulturwissenschaft. Am Beispiel einer Geschichte des sprachlichen Umbruchs im 20. Jahrhundert. In: Heidrun Kämper / Ludwig M. Eichinger (Hrsg.): Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache. Tübingen: Narr. S. 419-439. (= Studien zur deutschen Sprache, 40)
- Kaufmann, Michael / Wagner, Dorothea (Hrsg.) (2001): Drawing Graphs. Methods and Models. Berlin / Heidelberg / New York et al.: Springer.
- Konerding, Klaus-Peter (1993): Frames und lexikalisches Bedeutungswissen. Untersuchungen zur linguistischen Grundlegung einer Frametheorie und zu ihrer Anwendung in der Lexikographie. Tübingen: Niemeyer.
- Landgraf, Britta (2001): 3D Graph Drawing. In: Michael Kaufmann / Dorothea Wagner (Hrsg.): Drawing Graphs. Methods and Models. Berlin / Heidelberg / New York et al.: Springer. S. 172-192.
- Neidhardt, Friedhelm / Rucht, Dieter (1993): Auf dem Weg in die „Bewegungsgesellschaft„? In: Soziale Welt 44, S. 305-326.
- Nelson, Thomas E. / Clawson, Rosalee A. / Oxley, Zoe M. (1997): Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance. In: The American Political Science Review, Vol. 91, No. 3, S. 567-583.
- Perkuhn, Rainer/Belica, Cyril (2006): Korpuslinguistik – Das unbekannte Wesen. Oder Mythen über Korpora und Korpuslinguistik. In: Sprachreport 22 (1), S. 2-8.
- Rettie, Ruth (2004): Using Goffman's frameworks to explain presence and reality. In: 7th Annual International Workshop on Presence; 13-15 October 2004, Valencia, Spain. S. 117-124.
- Sarasin, Philipp (2001): Reizbare Maschinen Eine Geschichte des Körpers 1765-1914. Frankfurt am Main: Suhrkamp. (= stw 1524)

- Sarasin, Philipp (2004): „Anthrax“. Bioterror als Phantasma, Frankfurt am Main: Suhrkamp.
- Sarasin, Philipp / Berger, Silvia / Hänseler, Marianne / Spörri, Myriam (Hrsg.) (2007): Bakteriologie und Moderne. Studien zur Biopolitik des Unsichtbaren 1870-1920. Frankfurt am Main: Suhrkamp.
- Scharloth, Joachim / Noah Bubenhofer (2012): Datengeleitete Korpuspragmatik: Korpusvergleich als Methode der Stilanalyse. In: Ekkehard Felder / Marcus Müller / Friedemann Vogel (Hrsg.): Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen von Texten und Gesprächen. Berlin, New York: de Gruyter. S. 195-230.
- Scharloth, Joachim / Christian Gerber / Balthasar Glättli / Michel Studer / Noah Bubenhofer / Sarah Ebling / Saskia Vola (2010): Die Schweiz in der Krise: Korpuspragmatische Untersuchungen zur sprachlichen Konstruktion und Diffusion von Krisensemantiken. In: Aptum 2/2010. „Sprache(n) der Krise“ – International vergleichende Perspektiven. S. 99-120.
- Scharloth, Joachim (2005): Sprachnormen und Mentalitäten. Sprachbewusstseinsgeschichte in Deutschland im Zeitraum von 1766 und 1785. Tübingen: Niemeyer. (= RGL 255)
- Schiller, Anne / Teufel, Simone / Thielen, Christine (1995): Guidelines für das Tagging deutscher Textcorpora mit STTS. Stuttgart: [Working Paper] Universität Stuttgart, Institut für maschinelle Sprachverarbeitung, Tübingen: Seminar für Sprachwissenschaft.
- Schmid, Helmut (1994): Probabilistic part-of-speech tagging using decision trees. Stuttgart: [Working paper] Universität Stuttgart, Institut für maschinelle Sprachverarbeitung.
- Sinclair, John (1991): Corpus, concordance, collocation. Oxford: Oxford University Press.
- Spitzmüller, Jürgen / Warnke, Ingo H. (2008): Methoden und Methodologie der Diskurslinguistik – Grundlagen und Verfahren einer Sprachwissenschaft jenseits textueller Grenzen. In: Dies. (Hrsg.): Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene. Berlin/New York: de Gruyter. S. 3–54 (= Linguistik – Impulse & Tendenzen 31).
- Spitzmüller, Jürgen / Warnke, Ingo (2011): Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse. Berlin/New York: de Gruyter (= de Gruyter Studium).
- Steyer, Kathrin (2004): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: Dies. (Hg.): Wortverbindungen – mehr oder weniger fest. Berlin, New York: de Gruyter. S. 87–116. (= Institut für Deutsche Sprache. Jahrbuch 2003)
- Steyer, Kathrin / Brunner, Annalen (2009): Das UWV-Analysemodell. Eine korpusgesteuerte Methode zur linguistischen Systematisierung von Wortverbindungen. Mannheim: Institut für Deutsche Sprache (Online publizierte Arbeiten zur Linguistik OPAL 1). Elektronische Ressource. URL: <http://www.ids-mannheim.de/pub/laufend/opal/privat/opal09-1.html>
- Steyer, Kathrin / Lauer, Meike (2007): Corpus-Driven. Linguistische Interpretation von Kookkurrenzbeziehungen. In: Eichinger, Ludwig M./Kämper, Heidrun (Hgg.): Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache. Tübingen: Narr. S. 493–509. (= Studien zur deutschen Sprache 40)
- Tamassia, Roberto (Hrsg.) (in Druck): Handbook of Graph Drawing and Visualization. Boca Raton, London: Chapman & Hall/CRC Press.

- Teubert, Wolfgang / Čermáková, Anna (2007): Corpus linguistics. A short introduction. London: Continuum.
- Tognini-Bonelli, Elena (2001): Corpus linguistics at work. Amsterdam: Benjamins. (= Studies in corpus linguistics 6).
- Ziem, Alexander (2008): Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz. Berlin/New York: de Gruyter.